6-1-1997

Methods of digital classification accuracy assessment

Jeffrey R. Allen

Methods of Digital Classification Accuracy Assessment

M.S. Thesis

by

Jeffrey R. Allen

B.S. Rochester Institute of Technology
(1996)

Rochester Institute of Technology
Center for Imaging Science
Digital Imaging and Remote Sensing Laboratory

June 1997

Acknowledgments

I would like to acknowledge the following people for their contributions...

First, the efforts of all three members of my thesis committee. Your time and patience are greatly appreciated. I am particularly thankful to Dr. Navalgund Rao for kindling my interest in digital image processing, Mr. Rolando Raqueño for always taking time from his hectic schedule to deal with my insignificant problems, and Dr. John R. Schott for exposing me to the field of remote sensing, providing me with the opportunity to conduct this research, and helping shape my future.

I am grateful to the National Reconnaissance Office (NRO) for providing funding, without which this thesis would not have been possible.

Words cannot begin to express my appreciation towards my mother, Lorna. Her influence has been permanent, her support without bounds, and her love unconditional and absolute.

All of my family for their love and support including my sisters Kim and Maryann, brothers Tim and Scott, and the newest member to join our ranks, my adorable niece Rachel Allen. In addition, I would like to congratulate Scott and his fiancée Rebecca on their pending nuptials.

My girlfriend of nearly six years, Tracie Bonacci, for her love and understanding through years of schooling and beyond.

All of the professors I have had pleasure of knowing from the Center for Imaging Science and the College of Science at RIT. Your efforts are appreciated.

My friends from the DIRS lab, the Center, and Triangle fraternity for helping maintain my sanity (I think).

Mr. Stephen L. Schultz, unix wizard, for his programming assistance and Mr. Scott D. Brown, DIRSIG guru, for generating the required synthetic images.

All the support staff at the Center, especially, Mrs. Sue Chan for her academic planning.

Dr. Paul Wilson for his contributions toward the content of this thesis along with his recommendations and useful comments.

Lastly, all the unmentioned friends, colleagues, and study partners I have collected over the last 6 years at RIT.

Thank you all!

Dedication

This thesis is dedicated to my Father, Edward Oliver Allen Jr. It is difficult to be brief when I speak of him. His kind nature, endless patience, and limitless curiosity has shaped my personality and identity. I am well served by attempting to emulate him in every respect. My father's technical knowledge and mechanical expertise is inspirational to all who have known him and has made him my greatest mentor. I consider myself very lucky to have known such a dedicated father and ideal role model. Your impact is eternal, but your presence is missed.

APPROVAL OF M.S. THESIS

Methods of Digital Classification Accuracy Assessment

by
Jeffrey R. Allen
B.S. Rochester Institute of Technology
(1996)

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Imaging Science from the Center for Imaging Science, Rochester Institute of Technology.

June 1997

Signature of Author: ______________________________
Jeffrey R. Allen Social Security Number

Accepted by: ______________________________ June 6, 1997
Dr. Harvey Rhody, Coordinator, M.S. Degree Program Date

Center for Imaging Science
Rochester Institute of Technology
Rochester, New York

Certificate of Approval

Master of Science Degree Thesis

The Master of Science degree thesis of Jeffrey R. Allen has been examined and approved by the thesis committee as satisfactory for the thesis requirement for the Master of Science Degree.

Dr. John R. Schott, Primary Thesis Advisor

Dr. Navalgund Rao, Committee Member

Rolando Raqueño, Committee Member

6/3/97
Date

Center for Imaging Science
Rochester Institute of Technology
Rochester, New York

Thesis Release Permission Form

Thesis Title:

Methods of Digital Classification Accuracy Assessment

I, Jeffrey R. Allen, grant permission to the Wallace Memorial Library of the Rochester Institute of Technology to reproduce this thesis in whole or in part provided any reproduction will not be of commercial use or for profit.

Jeffrey R. Allen

6/3/97
Date

1. Abstract

Landcover classification of remotely sensed data has found many useful applications in industries such as forestry, agriculture, and defense. With the push toward end users, class maps are often incorporated directly into geographical information systems for use in solving large, complex problems. However, errors are inherent in the classification process. The importance of assessing the thematic accuracy of data derived from remote sensing platforms is universally recognized and has motivated much research. Classification accuracy assessment is often required to determine the "fitness of use" or suitability of a data set for a particular application. Failure to identify the magnitude of inaccuracies in classified data can result in errors cascading into subsequent exploitation and eventually result in false conclusions or flawed products. Many different techniques have been developed and utilized by the remote sensing community for performing thematic accuracy assessment. To date, no one procedure has been adopted as an industry-wide standard.

The purpose of this research was to evaluate the effectiveness and compare the results of several state-of-the-art assessment techniques. Synthetically generated imagery, along with real multispectral line scanner data, served as the baseline for the comparison. Synthetic imagery is uniquely suited for this task because the exact classification accuracy can be determined.

Table of Contents

List of Figures

List of Tables

2. Introduction

Digital image classification is one of the most common operations performed on remotely sensed data. Classification refers to a process where each pixel in an image is assigned to a certain category, known as a class. In the context of remote sensing, these classes usually correspond to types of ground cover. The result of classification is known as a class map. The term 'map' should not be confused with the cartographic meaning. A class map is digital raster data where digital counts (DC) correspond to class membership and spatial location corresponds to the same location as in the original image. Recent interest in the integration of remote sensing data into geographical information systems (GIS) has rekindled research and heightened interest in classification accuracy assessment (Janssen and Van der Wel, 1994). The push towards real world applications and the end user has further increased the need for reliable methods of accuracy assessment. Errors are introduced into classification when a pixel is misclassified by assigning it to the wrong class. The term pixel (picture element) is used to refer to the smallest element of the original and classified images. The original and classified images consist of a two dimensional array of pixels but the original image usually has an additional dimension of spectral data as well. Ideally, accuracy assessment would consist of comparing the class of all pixels in a classified image to their true class. In practice, accuracy assessment consists of comparing a small sampling of classified pixels to a set of data believed to be their true class. Over the years, many methods for accuracy assessment have been presented in remote sensing literature but no dominant standard has yet been adopted.

In this thesis, the current state-of-the-art accuracy assessment techniques are presented and a few unique adaptations are proposed, as well. These assessment techniques are then implemented on a series of baseline images. Three scenes are used for this purpose: the *tank* scene, the *desert* scene, and the *forest* scene. The first two images were acquired using airborne multispectral line scanners while the last image was synthetically generated. Details about these image sets can be found in §5.4. Classifier performance is affected by many real world imaging parameters. Two such parameters are image resolution and atmospheric visibility. The three scenes were degraded using these two parameters to create nine images. These nine images were useful because, after classification, they provided a complete range of classification accuracy which was needed to thoroughly compare the various assessment techniques. In addition, they were also used to quantitatively measure the effect of the stressing parameters on classification accuracy. Three difference classifiers were used to produce the requisite class maps: the Gaussian Maximum Likelihood (GML) using parametric multivariate statistics, the Fuzzy ARTMAP neural network utilizing a fuzzy logic set, and Mystic™ a new classifier using mathematical rules, optimized by a genetic algorithm, to segment classes. These three classifiers, described in §5.3, were selectively used on the nine images to generate twenty three class maps. All of these class maps then underwent accuracy assessment based on a variety of reference data sources. The result was one hundred and nineteen confusion matrices and several corresponding accuracy metrics for each. For the exact combination of image, classifier, and reference data the reader is referred to the experimental matrices in §6.1. The research of the thesis is divided into five major thrusts: obtaining reference data, accuracy metrics, parameters stressing classifier performance, correction for biased reference data, and the relative performance of the rule based classifier. Each of these topics is discussed in greater detail below.

2.1 Collecting Reference Data

The process of classification accuracy assessment can be grouped into two distinct steps. In the first step, the class map is spot checked against reference data. The second step involves calculating a meaningful metric from the data collected in the first step. Collecting reference data will be introduced in this section, and accuracy metrics in the next. Classification accuracy assessment is presented, in detail, in §7.2. Reference data is a group of pixels which belong to known classes that are used to estimate the accuracy of the entire map. Several methods of obtaining reference data is presented in §7.3. When assessing classifier performance, reference data is compared against the class map to build a confusion matrix. A confusion matrix is a contingency table, often used in categorical data analysis, usually with reference data along the columns and class map data along the rows. In each element, along the row and column of the confusion matrix, the corresponding number of pixels that fall into both categories is posted. Any discrepancy between reference data and classified data is considered a classification error. The difficulty lies in obtaining reference data, sometimes known as verified, identified, known, or truth data, which is representative of the entire scene.

Determining the exact accuracy of a class map is impossible in almost all circumstances. For certain, it is impractical in all cases involving real imagery. There are, however, several widely accepted methods for formulating a reasonably close approximation to the true accuracy. A proper estimate will also include the corresponding confidence interval. When selecting a method for accuracy assessment, there is a trade off between cost and accuracy. The cost of accuracy assessment includes many factors such as labor, physical resources, time, travel, and others. The largest cost of assessment is incurred obtaining the reference data. Less robust methods result in less accurate approximations of accuracy with large confidence intervals but at a lower cost. High quality assessment procedures are more accurate, but also more expensive and time consuming. Each project must find the balance point between cost and acceptable fidelity. Many accuracy assessment techniques introduce bias into their estimations. Bias is the systematic error resulting from consistent over or under estimation of the true class map accuracy. Optimistic and conservative bias will be discussed in §8.1 and §8.2 respectively. The source of this distortion can often be traced to the source of the reference data.

For this thesis, real and synthetic imagery will be employed for comparing different sampling techniques. In the context of classification accuracy assessment, sampling refers to selecting certain pixels, or groups of pixels, and determining which class they truly belong in. Synthetic imagery is of particular interest because sampling is not necessary since the exact class membership of each pixel is known. This *a priori* knowledge will permit an unbiased, quantitative evaluation of the popular sampling techniques. The use of synthetic imagery in this research is explained in §6.3.

Over the years, several sampling techniques have been employed by the remote sensing community for this purpose. However, each sampling technique has corresponding advantages and disadvantages. In §9.1 the results of the analysis of the effect of reference data source on the reported accuracy metric will be detailed.

2.2 Accuracy Representation

Once reference data is used to create a confusion matrix, it is often desirable to reduce the matrix into a single, meaningful index of accuracy. This single metric, usually expressed as a coefficient between zero and one, estimates the true average map accuracy or classifier performance. Many different accuracy metrics have been introduced to compensate for the fact that the estimate is being made on less than complete information. Other metrics, ideal for measuring classifier performance rather than class map accuracy, correct for the proportion of pixels properly classified only by chance. It is important to keep in mind the method used to generate the confusion matrix when selecting this metric. The most often quoted metrics are the Simple accuracy, Weighted accuracy, Kappa coefficient, Brennan and Prediger's Kappa, and the Tau coefficient which will be introduced in §7.4. This lack of a standard has created difficulties in comparing different class maps. A conversion from one metric to another cannot be made because they also depend on the marginal distribution of the confusion matrix in most cases. In this thesis, a comparison between different accuracy metrics will be made noting the advantages and disadvantages of each. The appropriate confidence interval, accounting for uncertainty from all sources, will be reported along with this metric. This will be accomplished with highly characterized real imagery and computer generated synthetic images.

The metrics will be evaluated on class maps generated with the GML classifier, the Fuzzy ARTMAP classifier, and Mystic™, a rule based classifier. Supervised classification algorithms and uncorrected images will be employed in this study because they are most commonly used by the remote sensing community. Methods for accuracy assessment and accuracy metrics are normally considered completely independent of the classification technique utilized. However, because classifiers may exhibit different degrees of spatial correlation of errors, three different classifiers will be used to ensure universal applicability of the results. The baseline images will also contain a variety of land covers to avoid correlation in the final results. In §9.2, the results of the analysis and comparisons between the accuracy metrics are covered.

2.3 Factors Degrading Classifier Performance

In addition to image content and the quality of training data, the accuracy of image classification is a function of several real world imaging parameters. A discussion of several factors effecting classification accuracy is contained in §7.1. However, only two significant factors were examined as part of this research. The first, image resolution, was examined using the *desert* scene. The second factor, atmospheric visibility, was analyzed using the *forest* and the *tank* scenes. Both stressing parameters were simulated using the procedure outlined in §6.4. To determine the extent of the effect on classification accuracy, several accuracy assessment procedures were employed. Finally, the effect of these stressing parameters on the classification accuracy of the baseline images is presented in §9.3.

2.4 Correcting for Reference Bias

Quality reference data to be used for classification accuracy assessment is often difficult, time consuming, and costly to obtain. Often analysts utilized user selected reference as a quick, low-cost altemative to rigorous random verification. However, user selected reference data almost always suffers from overly optimistic bias. User selected reference also has another problem. In general, its marginal distribution in the resulting confusion matrix does not accurately approximate the true class probability distribution. In §8.3, a method is proposed to correct for this shortcoming. This process is classed confusion matrix marginal distribution scaling by post priori probabilities. It is used in this thesis to adjust the confusion matrices of all three scenes constructed using independent reference data. The accuracies resulting from scaled matrices are then compared to the unscaled and true accuracies. These results are presented in §9.4.

2.5 Relative Classifier Performance

The last area of research is the performance of the Mystic™ classifier relative to the other two, more traditional, classifiers. Mystic™ is a new, rule based classifier. It uses a genetic algorithm to optimize the parameters of the rules to obtain the highest classification accuracy possible. The Mystic™ classifier, along with the GML, and Fuzzy ARTMAP are developed in §5.3. To this point in time, the accuracy and properties of the Mystic™ classifier are relatively untested. It however appears to be a unique classifier with promising potential. While a major thrust of this thesis is not a comparison between classifiers, preliminary results and suggestions for the Mystic™ classifier are presented in §9.5.

3. Objectives

The objective of this thesis is several fold. First, the current techniques of classification accuracy assessment, also known as classification validation, will be presented. One objective will be to develop a common formalism and taxonomy of accuracy assessment. Many independent researchers have presented results on classification accuracy assessment. Several contrasting approaches have been given, as well. Many of these papers have used different terminology even when referring to the same phenomenology because no standards yet exist. Ideally, this thesis will serve as a compendium of classification validation by providing a common source of research results drawn from years of remote sensing literature.

3.1 Analysis of Accuracy Assessment

In this project several difference sampling schemes will be employed. The accuracy of these schemes will be determined using synthetic reference or more rigorous sampling. Corrections for reference data which poorly estimates the true class probability distributions will also be made. Accuracy metrics will be evaluated and compared in a similar method. In addition, it will be determined if each metric is accurately estimating the quantity it is supposed to be measuring.

3.2 Application of Accuracy Assessment

As part of this project, a database with a significant number of confusion matrices has been generated. In this thesis, the purpose of these matrices was to analyze accuracy metrics, reference data sources, classifier performance, and the effect of stressing situations. Ideally, these same confusion matrices could be used in the future to analyze other factors.

When analyzing the effect of a parameter on classification accuracy, it is often difficult to separate the effect of the desired parameters from the effect of the assessment procedure. The quantitative assessment of stressing parameters will inevitably include bias introduced by the method of assessment. Different assessment methods will result in differing values. This is because it is difficult to obtain an accurate or precise accuracy assessment. For this reason, the evaluation of stressing parameters and classifier performance in this thesis is combined with the analysis of assessment procedures.

4. Work Statement / Deliverables

Statement of Work

» Designate and optimize a common training set to be used by all classifiers.

» Classify candidate imagery with GML, Fuzzy ARTMAP, and rule based genetic algorithm (Mystic™) classifiers using common training data.

» Obtain reference data from dependent, independent, random, and synthetic sources.

» Generate confusion matrices from reference data and evaluate classification accuracy using Simple, Weighted, Kappa, B&P's Kappa, and single class accuracy coefficients.

» Generate and analyze confusion matrices made from user selected reference which have been scaled to match *post priori* distributions.

» Utilize synthetic imagery to identify most precise and efficient method of classification accuracy assessment.

» Analyze effect of stressing parameters on classification accuracy and relative effectiveness of Mystic™ classifier.

List of Deliverables

» Program for converting ENVI™ training data to Mystic™ and AVS™ format.

» A Mathematica library for generating confusion matrices from dependent, independent, random point and synthetic data sources.

» A Mathematica library for evaluating Simple, Weighted, Kappa, B&P's Kappa, and single class accuracy coefficients with confidence intervals.

» A written database containing confusion matrices for all classified images based on user selected, random, and synthetic reference.

» A written document detailing state of the art classification accuracy assessment techniques employing a common vocabulary and formalism.

» A written document containing suggestions for minimizing bias and increasing precision of classification accuracy assessment.

» A written document covering the theory, background, approach, and results of the study.

5. Background

5.1 Utility of Classification

As mentioned previously, digital image classification is one of the most important processes when preparing remotely sensed data for use in applications or research. Different users sometimes refer to image classification as class segmentation, categorization, or landcover determination. A variety of users have found classification of satellite and aerial images a cost effective solution to challenging large scale problems. However, the synoptic view, high availability, and frequent overflights has made satellite imagery the preferred, low-cost data source of many users. Classified images are known by several names including class maps, thematic maps, product maps, land-use maps, and landcover maps. Classification can be used to determine the land cover, constituent material type, or object class of each pixel in an image acquired at great distances.

The environmental community has made wide use of classification as a tool when studying large areas of isolated environments. Data is often collected over time to monitor environmental change such as deforestation and changes in wetlands. Classification has proven to be an invaluable aid in the mapping of wildlands and in drafting inventories of isolated locations (Fitzpatrick-Lins, 1980; Senseman *et al*, 1995; Rutchey and Vilcheck, 1994).

National governments have been the largest user of classification on remote sensing data. It is often used for surveying and monitoring vast natural resources (Bauer *et al*, 1994). For example, classification has helped optimize water usage in developing countries (Nageswara Rao and Mohankumar, 1994). Governments have also been successful in predicting crop failure and avoiding famine in developing nations by preparing relief aid in advance. Many other applications such as urban planning and defense related uses have also benefited.

The commercial sector has also found many useful applications. Segmented images often aid in oil exploration, identification of mineral deposits in remote locations, populating GIS databases and cartography. Class maps of crops have been used in many ways to optimize agriculture. Crop yields can be maximized by determining where and when it is best to plant, fertilize, irrigate, or harvest. It has been used by the logging industry to identify and manage forest resources. Even large brokerage houses have utilized classification of remotely sensed images for predicting commodity and future prices by monitoring crop health and measuring biomass. Many of these applications base important decisions on evidence uncovered by image classification. This underscores the need for high quality class maps where the accuracy and confidence intervals are known.

Classification is sometimes used as a preprocessor to further digital image processing. For example, class maps can be used for atmospheric calibration or emissivity determination in thermal studies. For these applications especially, class maps must be of high accuracy to ensure excessive error is not propagated to further processing steps. Classification is commonly used in image exploitation as an analyst's tool. It reduces the dimensionality of data with little or no loss of critical information which in turn aids in human assimilation (Harsanyi, 1994).

5.2 Motivations for Accuracy Assessment

There are several types of error introduced into remote sensing data. Other than radiometric, there are two major types of error that are of concern. The first, positional error, refers to the improper relative location of a pixel in a scene when compared to the original scene geometry. Positional accuracy is often measured in root mean square (RMS) units and corrected for using one of the methods of image registration or rectification. The second, thematic accuracy, is the focus of this thesis. It refers to errors in ancillary data associated with a certain pixel such as class membership. The focus of this thesis is thematic accuracy of classified images but positional accuracy affects the measurement of thematic accuracy. Positional error when collecting reference data from a second registered image will result in underestimated thematic accuracy. In addition, class map pixels with positional error will no longer correspond to proper relative location on the ground. In this thesis, accuracy, unless noted otherwise, will be referring to the thematic accuracy of a class map.

Accurate data is critical to all of the applications mentioned above. Classified images are of no use if they contain excessive amounts of error, therefore, the validation of classified data is paramount. The amount of tolerable error is specific to each application but the need for accuracy assessment is consistent. The thematic accuracy is often the deciding factor in determining whether a class map is appropriate for a study.

Precise accuracy assessment is needed to determine the effectiveness of different classifiers. Continuing research into new, more robust classifiers requires effective methods for measuring their performance. Vigorous accuracy assessments can also point out flaws in existing classifiers and lead to improvements. Assessment has also been used to facilitate studies to determine how imaging parameters such as view angle, time of day, spatial resolutions and even the sensor used affect the final classification accuracy.

5.3 Classification Algorithms

There are two distinct techniques of image classification. The first type is unsupervised classifiers such as k-means (Duda and Hart, 1973) and ISODATA (Tou and Gonzalez, 1974) algorithms. These routines are highly automated and require only one input from the user, the number of desired class categories. The categories segmented by these methods may or may not correspond to classes which may be desirable for the user. However, unsupervised classifiers are often used as a quick first run to determine how separable desired classes might be. They are also often used to provide pure training data to the second type of classifier. The second and more popular type of classification are the supervised algorithms. Supervised classification routines require prototype training data from the user.

Training data are sample pixels, along with thematic labels, which belong to the classes which the user wishes to segment. Once they have been selected, the entire spectral vector of each pixel is used by the classifier. The gathering of training data is a subjective, man-in-the-loop process, which has a large bearing on the ultimate classification accuracy. Supervised classification training data is usually identified by an image analyst using one of two techniques. With the first and most common method the analyst interactively selects solid polygons over areas of an image which are believed to contain only the desired class. The second technique requires the image analyst to only select a single point in the center of a homogeneous area of the image representing the desired image class. This single point is then used as a seed for an unsupervised classifier, such as the fuzzy k-means clustering algorithm, which extrapolates to select spectrally and spatially near image data to be used as training data. Both supervised training methods require the analyst to determine the number of desired classes, and select at least one region per category.

After the classifier is trained or developed with the training data, the supervised classifier proceeds to assign thematic labels to the pixels in the image. Most classifiers allow pixels to remain undefined which do not fit well into any of the established categories. In this case, the user must supply a membership coefficient threshold that must be exceeded for any given pixel to be classified. Undefined pixels will not increase or decrease measured accuracy because they are not included in confusion tables. Three supervised classifiers will be used in this project. They also are all 'per-pixel' classifiers. This means they assign pixels individually to classes based only on the spectral signature from that pixel, with no regard to the surrounding pixels. Other classification routines use a statistical measure of the local neighborhood to help assign pixels. This can help classifiers by recognizing the texture of a class even when the spectral signature of that class would not help segment it.

5.3.1 Gaussian Maximum Likelihood

The Gaussian Maximum Likelihood (GML) classifier is the most popular of all classifiers. The GML is a supervised classification algorithm which employs Bayesian probability theory to select the class to which a pixel most likely belongs. This is accomplished by segmenting feature space with n-dimensional clouds called hyperellipsoids. If the statistical assumptions set forth by this method are valid for a given data set, the resulting classification will minimize overall classification error. Because this classifier is so widely accepted and theoretically understood, it is often used as a benchmark for comparisons against new classifiers.

The subsequent derivation of the GML classification routine follows closely with that of Schott (1997). The GML classifier is most readily derived and visualized by considering a single band, gray scale image. This treatment of the univariate case is then able to be scaled to the multivariate case with the appropriate number of spectral channels. Using Bayesian probability theory, the *a posteriori* probability [$p(i|DC)$], is the probability that a pixel with an observed digital count of DC will belong to class *i*.

$$p(i|DC) = \frac{p(DC|i)\, p(i)}{p(DC)} \tag{5-1}$$

The *a priori* probability, p(i), is the probability that any class *i* will be observed. In other words, this term is the proportion of pixels which belong in the class *i*. The chance that a particular digital count DC will be observed within a certain class is given by $p(DC|i)$. This value is evaluated by the GML classifier using Equation 5-2 for all values of DC and *i* based on the training data supplied by the user. A few years ago the computer storage requirement of this calculation was significant when dealing with multispectral and especially with hyperspectral imagery. Today, with modern computers, this same amount of storage is insignificant. Implicit in this treatment is the assumption that the digital counts within any given class have a Gaussian distribution. This is because the spectral distribution of classes are only represented by their means and standard deviations.

$$p(DC|i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\left(\frac{DC-\overline{DC}_i}{2\sigma_i^2}\right)^2} \tag{5-2}$$

Where: i is the class,
DC is the digital count of a pixel,
$\overline{DC}_i$ is the average DC of the class i and,
σ_i is the standard deviation of the class i.

The term p(DC) is the probability of any digital count occurring, otherwise known as the image wide normalized histogram. This function is the same for all classes and simply scales the resulting *a posteriori* probability. If the p(DC) term is dropped from the GML classifier, it will have no effect on the results. The goal is to find the class, i, with the highest probability not the value of the absolute probability. The rank ordering is maintained with the following simplified equation:

$$p(i|DC) \cong p(DC|i)\ p(i) \tag{5-3}$$

Bayes decision function is then defined to be the GML discrimination metric, D_i', by substituting Equation 5-2 into Equation 5-3. The GML discriminate metric (Equation 5-4) is the value by which class membership will be decided on a pixel by pixel basis.

$$D_i' = p(DC|i)\ p(i) = \frac{p(i)}{\sqrt{2\pi\sigma_i^2}} e^{-\left(\frac{DC-\overline{DC}_i}{2\sigma_i^2}\right)^2} \tag{5-4}$$

This metric can be further simplified by taking the logarithm of D_i'.

$$D_i'' = \ln[D_i'] = \ln[p(i)] - \frac{1}{2}\ln[2\pi] - \ln[\sigma_i] - \frac{(DC-\overline{DC}_i)^2}{2\sigma_i^2} \tag{5-5}$$

Finally adding a constant to Equation 5-5, we have arrived at the final GML discriminant shown as Equation 5-6. Neither taking the logarithm nor adding a constant will change the rank ordering determined by the discriminant.

$$D_i = \ln[p(i)] - \ln[\sigma_i] - \frac{(DC-\overline{DC}_i)^2}{2\sigma_i^2} \qquad (5\text{-}6)$$

At each pixel in the image, the discriminant, D_i, is evaluated for all classes, i. The class with the highest value is selected as the class of that pixel. In many implementations, the user is allowed to select a probability threshold which must be exceeded before a pixel can be assigned to a class. Pixels that are not assigned to a class are left as undefined in the final class map.

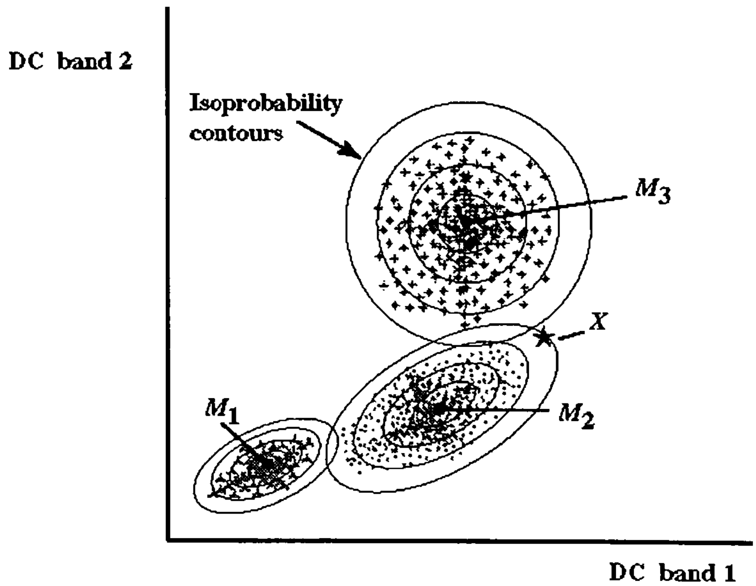

Figure 5-1 GML Classification of a Two Band Image

While the GML classifier has been derived thus far assuming a one band image is being classified, this is rarely the case in practice. Images which are normally classified have multiple bands. In this case each pixel is an n-dimensional spectral vector. When classifying multidimensional imagery, the algorithm is the same except scalar mathematics is replaced with vector mathematics. For example, take an n-dimensional spectral vector of a pixel $\bar{x}$ and a m-dimensional vector $\bar{w}$ of the target classification classes where n is the number of image bands and m is the number of target classes.

$$\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{pmatrix} \qquad \bar{w} = \begin{pmatrix} w_1 \\ w_2 \\ . \\ . \\ . \\ w_m \end{pmatrix} \tag{5-7}$$

In this case, Equation 5-1 would need to be transformed to its vector equivalent form shown by Equation 5-8. The same is true for the rest of the calculations in the GML classifier.

$$\mathrm{p}(w_i|\bar{\mathrm{x}}) = \frac{\mathrm{p}(\bar{\mathrm{x}}|w_i)\ \mathrm{p}(w_i)}{\mathrm{p}(\bar{\mathrm{x}})} \tag{5-8}$$

Figure 5-1 has been provided to aid in visualizing the classification of a two band image containing three distinct classes. The three classes are centered about their respective multivariate means $\boldsymbol{M_1}$, $\boldsymbol{M_2}$, and $\boldsymbol{M_3}$. The concentric ellipsoids centered about these means represent iso-contour intervals of equal class membership probability or GML discriminate value. The distribution of pixels has both a mean in band one and in band two. However, as seen in Figure 5-1, the distribution can take on a diagonal character as well. This is due to correlation in digital counts of classes in multiband images. The multivariate statistical approach taken by the GML classifier accounts for the shape of this type of distribution with a covariance matrix. This ability of the GML classifier results in higher classification accuracy than similar classifiers such as the parallelepiped classifier which lacks this ability.

Unlike the Fuzzy ARTMAP and the Rule Based Genetic Algorithm which will be discussed below, the GML is a traditional, parametric classifier. It uses multivariate statistics and makes decisions using class orientation and spectral extent information contained in the mean vector and variance-covariance matrix. This parametric model minimizes effects of noisy or outlying training data due to its averaging properties. This advantage is moderated by the fact that image data which varies greatly from normal can be problematic. It is commonly noted that GML performance maps best to visual interpretation when compared to non-parametric classifiers. The Environment for Visualizing Images (ENVI™) software package was selected for its GML implementation for use in this thesis.

5.3.2 Fuzzy ARTMAP

In recent years, several neural-network type architectures have been implemented to classify images. The interest in neural-networks for use in classifiers is due to their ability to learn and remain flexible. Their rule for deciding in which category to classify a pixel will change and adapt from region to region in an attempt to make optimal decisions. Traditional neural-network classifiers have two primary disadvantages. First, neural-networks use traditional logic which allows for only crisp set, binary decisions. Secondly, conventional networks have required excessive amount of training cycles, or epochs. The fuzzy ARTMAP supervised classifier, developed by Grossberg and Carpenter at Boston University in 1991, overcomes both these limitations. It combines a fast learning neural-networks architecture with fuzzy logic decision making. Underlying principles of the network's operation are based on modeling of the human eye-brain system. The fuzzy ARTMAP architecture's ability to learn and adapt make it well suited to the classification of remotely sensed images. It will be one of the supervised classifiers utilized in this thesis as implemented by Nessmiller (1995) in Advanced Visualization System (AVS) environment.

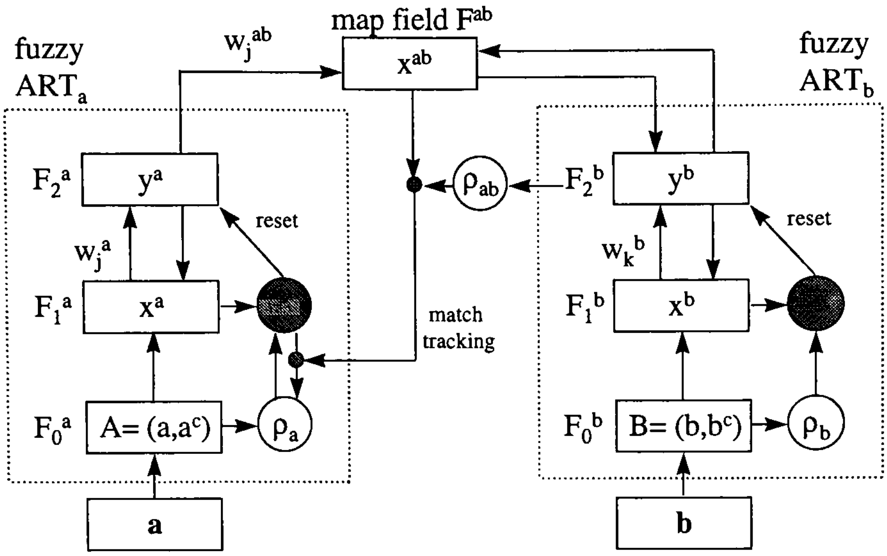

Figure 5-2 Fuzzy ARTMAP Architecture

(Nessmiller, 1995)

The Fuzzy ARTMAP does not have a classical mathematical derivation as does the GML classifier. The fuzzy ARTMAP classifier consists of an advanced neural network known as the Adaptive Resonance Theory MAPping (ARTMAP) combined with fuzzy logic algorithms. The architecture of the fuzzy ARTMAP, shown in Figure 5-2, will help briefly describe its operation as outlined by Nessmiller (1995). The fuzzy ARTMAP classifier consists of two Adaptive Resonance Theory (ART) neural networks, labeled ART_a and ART_b. The ART's are unsupervised classifiers by themselves. Two ART's can be combined to form a supervised classifier known as an ARTMAP. An ARTMAP can be modified to incorporate fuzzy logic which then forms the fuzzy ARTMAP classifier. The first step, as with any supervised classifier, is to supply training pixels. The spectral vector from a training pixel is supplied at **a** and the corresponding class label is supplied at **b**. The intensity values of each of the N spectral bands in the training data must first, however, be normalized to values between 0 and 1. Next, both inputs undergo a calculation called *complement coding* at the *preprocessing fields* F_o^a and F_o^b. Complement coding simply subtracts the normalized intensity value from 1. At F_o^b, the label is encoded with a binary designator which will be used by the network to specify the class categories. Next comes the long term memory of the fuzzy ARTMAP which consists of *the activity vector* F_1 and the *classification vector* F_2.

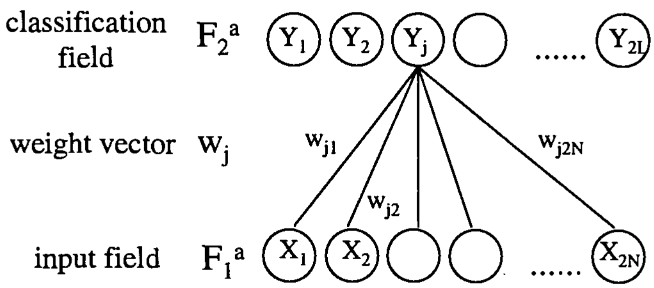

Figure 5-3 Weight Vector Operation

(Nessmiller, 1995)

Before training, all the weight vectors are set to unity. The goal of this network is to find the strongest connection of the weight factor between the input field and the classification field. However, before the classification can be considered acceptable it must meet or exceed the vigilance parameter. The vigilance parameter, ρ, is a certainty threshold which must be exceeded in order to classify a pixel in a given class. The higher the value, the more certain the classifier must be. This is an example of fuzzy logic within the fuzzy ARTMAP.

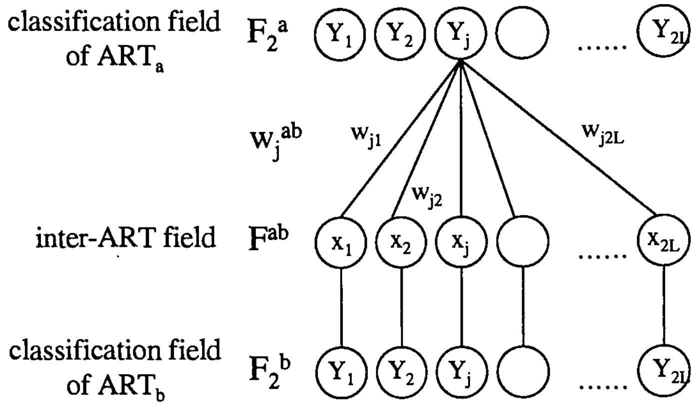

Figure 5-4 Inter-ART Field Operation

(Nessmiller, 1995)

The last step is the inter-ART field, represented by F^{ab}, which couples the two ART's together. The inter-ART field has two purposes. First, it maps the classification from ART_a to the classification output of ART_b. Secondly, it realizes the match tracking rule. When there is a mismatch during training between the output of ART_a and the correct classification of ART_b, *match tracking* occurs. Compared to other image classifiers, the fuzzy ARTMAP tends to be mathematically complex and computationally intensive. For a more rigorous development of the Fuzzy ARTMAP classifier, the reader may consult Nessmiller (1995) or Carpenter, *et al* (1991).

The fuzzy ARTMAP is a non-parametric classifier so it makes no assumption of normality as the GML classifier does. However, like other non-parametric classifiers, experience has indicated that it tends to be extremely sensitive to biased training sets and noisy data points. For this reason it requires a highly homogenous training set. This property is important to remember when selecting training regions. Therefore, the criterion is very different when selecting training sets for the fuzzy ARTMAP when compared to the GML classifier. Nevertheless, when supplied with robust training data it can provide very high classification accuracy.

5.3.3 *Rule Based Genetic Algorithm*

Mystic™ is a classifier, termed terrain categorization (TERCAT), which uses logical rules to assign image pixels to their respective classes. It has been implemented within the MATRIX™ environment. Rules can be powerful and flexible methods for associating an observed pixel with a specific class. Mystic™'s reliance on rules rather than statistics allows the classifier to make no assumption of normality. Therefore, this type of non-traditional classifier does not make the same errors that other traditional classifiers, such as the GML, make by erroneously assuming target reflectance is distributed in a Gaussian manner. Rules are simply a logical statement which selects some pixels and rejects others. A sample rule (Equation 5-9) is provided to illustrate the classification process. Parameters within each rule are optimized in such a way the that the rules function in the best manner possible on the supplied training data. The measure of how well a specific rule functions is based on its performance during the optimization process were it is used against the training data, where the 'true' class in known. This measure for a given rule is called a reward function and is calculated by applying the rule to all pixels in the training set and finding the number of correctly classified pixels. The more pixels properly classified, the higher the reward value for that combination of variables. In other words, the dependent set accuracy assessment is used as feedback into the classifier. Obviously, assessing the accuracy with this same data set will result in an overly optimistic accuracy estimate. The enormous amount of parameter combinations allowed by even simple rules necessitates the use of an advanced optimization algorithm. Attempting to test each combination is precluded due to practicalities of time constraints on any current or foreseeable computer. Recent developments of sophisticated optimization methods have made rule based classifiers practical. Mystic™ uses a genetic algorithm (GA). Genetic algorithms were first defined by Friedberg (1958) and are named so due to their search technique which is analogous to a biological system. The GA is applied to Mystic™ by allowing the fittest rules, those with the highest reward function, to continue to the next generation. Even with this advanced optimization technique, the Mystic™ classifier is extremely computation intensive.

One of the simplest and most successful rules is given below. This typical rule is called the *one band threshold*. Once selected, this rule would be optimized by Mystic™ on the entire supervised training set provided by the user. The reward function for the set of optimization variables i,j,k is the number of pixels correctly classified when the prototype rule is applied to the training set. The set of optimization variables with the highest reward function is then selected and used with the rule to classify the entire image. Theoretically, once a rule is optimized it can be applied to other, similar data sets.

If:	$j \le b_i \le j + k$		(5-9)
	Where:	b_i is the DC in the i^{th} band and, i,j,k are variables optimized by the GA.	
Then:	b_i belongs to class associated with i, j, and k.		

Mystic™ requires that the user select the rules which will be used to identify pixels in each of the classes. Mystic™ is packaged with 6 predefined rules and allowances are made for user defined rules. A different rule can be used to identify each class but only one rule is allowed within each class. For example, different rules can be used to assign pixels to class A or class B. But only one rule can assign pixels to class A and only one rule can classify pixels as class B. The Mystic™ algorithm uses the GA to optimize the parameters of each rule, but not which rule is used. Currently, the Mystic™ classifiers are very simple and utilize only spectral information of each pixel. All rules are based on the DC in the bands of one pixel without regard to the neighboring pixels. Neglecting the surrounding pixels fails to utilize any of the spectral information of a scene which could be useful in classifying pixels but narrows an already vast search space.

5.4 Image Data Sets

Three different scenes were selected to be used in this study. Of these images, one was synthetically generated on a computer while the rest were acquired using real airborne sensors. These particular images were selected because they represent a wide sampling of terrain, phenomenology, and content. The M7 and Daedalus sensors used to acquire these multispectral images are of particular interest because of their combination of high spectral and spatial resolution. This combination has a great potential for generating images which can be classified to a high degree of accuracy and precision. The images used in this project where taken in the visible (VIS) to short-wave infrared (SWIR) spectral region of the electromagnetic spectrum. Bands longer than this, if any, were eliminated to avoid thermal photon contributions. Thermal bands are often avoided when classifying images because these bands have low day-to-day correlation. This attribute is not desirable because it makes training data collected from one image not applicable to images acquired on subsequent days. Portions around the perimeter of two images have been removed because they exhibited erroneous sensor effects. These portions where not classified and did not contribute to accuracy assessment. The images consisted of raw digital counts.

Nearly any study involving different image classification algorithms will utilize the GML classifier. The GML classifier has consistently demonstrated high classification accuracy and frequently is used as a baseline for comparisons of new classifiers. It was selected for use in this study for these reasons. However, the non-parametric nature of the Mystic™ classifier differs significantly from the GML. The non-parametric fuzzy ARTMAP classifier was chosen because it utilized a equally nontraditional approach as Mystic™.

5.4.1 Tank Scene

The first image (Figure 5-5), which will be called the *tank* scene, was acquired as part of the Southern Rainbow collection by Environmental Research Institute of Michigan (ERIM). It was captured at 8-bits per pixels using the 16 band M7 aerial line scanner. Band number 16, the thermal band, was removed and not used in this study. The bandpasses for the remaining bands are listed Table 5-1. This image in particular was selected for its diversity of content. In addition to forest, brush, and exposed soils, the scene contains a variety of man-made objects. The scene derived its name from the fact that several military vehicles, including tanks, are camouflaged throughout the image. During classification, all vehicles were categorized into one metal class. To reduce classification error and produce a useful class map, 9 classes were needed to categorize this image compared to approximately 5 for other scenes. This scene was imaged as part of a well organized collection and is therefore highly characterized. Many ground photos are available for building accurate reference data sets.

Table 5-1 Southern Rainbow Bandpasses

M-7 Band	Bandpass (μm)	"Color"
1	0.45 - 0.47	
2	0.48 - 0.50	Blue
3	0.51 - 0.55	
4	0.55 - 0.60	Green
5	0.60 - 0.64	
6	0.63 - 0.68	Red
7	0.68 - 0.75	
8	0.71 - 0.81	Near IR
9	0.81 - 0.92	
10	1.02 - 1.11	
11	1.21 - 1.30	
12	1.53 - 1.64	
13	1.54 - 1.75	
14	2.08 - 2.20	
15	2.08 - 2.37	

Figure 5-5 Southern Rainbow Tank Scene

5.4.2 Desert Scene

The *desert* scene (Figure 5-6) was acquired as part of the Western Rainbow, Joint Camouflage Concealment and Deception (JCCD) field collection using the Daedalus airborne sensor. The site of this scene is the Yuma proving grounds. The original GIFOV of the scene was one meter, but the image was also degraded to two and four meter resolutions for use in this study. The scene consists of mostly desert pavement (or desert varnish) but notable features have been expanded for illustration purposes in Figure 5-6. The thermal bands have been removed again and the edges which exhibited severe geometric distortion have be masked out. The collection was well documented and many ground photographs are available for verifying the land cover. The image was captured at 8 bits per pixel for each of the 10 final bands.

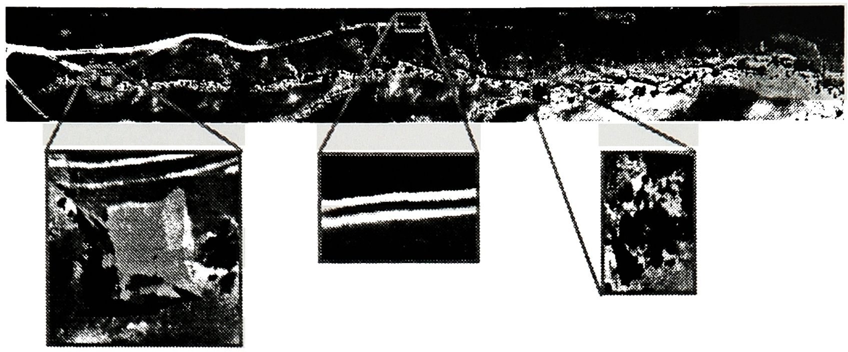

Figure 5-6 Western Rainbow Desert Scene

Table 5-2 Western Rainbow Bandpasses

M-7 Band	Bandpass (μm)	"Color"
1	0.405-0.455	Blue
2	0.435-0.535	
3	0.500-0.625	Green
4	0.570-0.650	
5	0.595-0.720	Red
6	0.645-0.790	
7	0.700-0.955	Near IR
8	0.785-1.070	
9	1.495-1.835	
10	2.011-2.560	

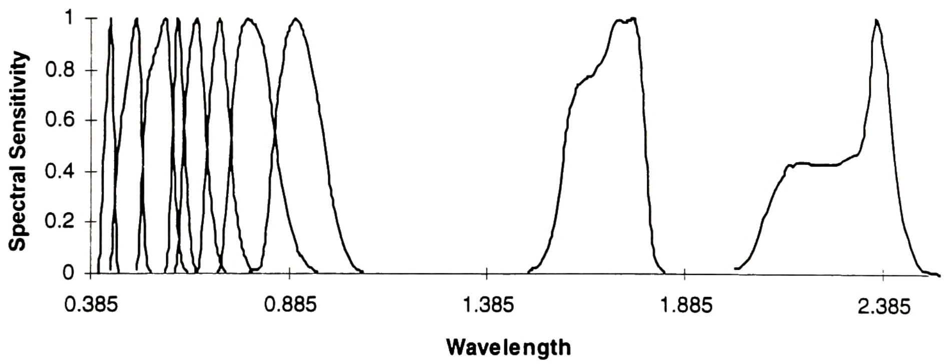

Figure 5-7 Bandpasses of Daedalus Sensor

5.4.3 Forest Scene

The *forest* scene (Figure 5-8) is the final image. Unlike the first two scenes, which were imaged with real airborne sensors, this image was generated synthetically with the Digital Imaging and Remote Sensing Image Generation (DIRSIG) model. The bandpasses (Table 5-3) simulate that of the M7 line scanner. The radiance field generated by DIRSIG was convolved with a 3x3 equal weighted kernel, resampled to one third of the original size using cubic convolution, and quantized to 8 bits per pixel for each of the 15 bands. Convolution was necessary because the radiance field pixels are spectrally pure but the convolution results contain mixed pixels, as is the case in real images. Three versions of the synthetic scene were generated. These images had LOWTRAN atmospheric visibilities of 23km, 7km, and 5 kilometers. For further details about synthetic images generated by DIRSIG, the reader is referred to *DIRSIG, Digital Imaging and Remote Sensing Image Generation, Description, Enhancement, and Validation* (Schott *et al*, 1993).

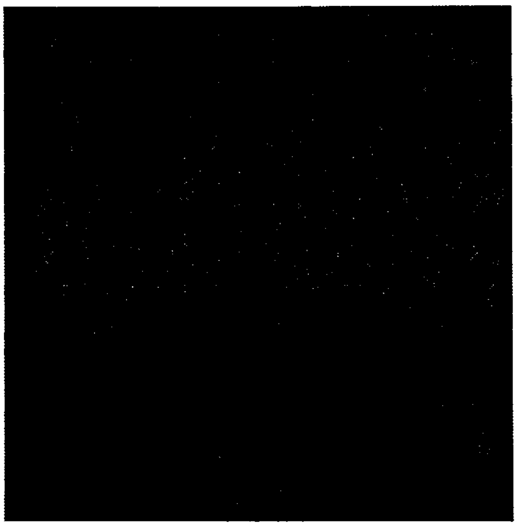

Figure 5-8 Synthetic Forest Scene

Table 5-3 DIRSIG Scene Bandpasses

Synthetic Band	Bandpass (μm)	"Color"
1	0.44-0.50	Red
2	0.46-0.53	
3	0.49-0.57	Green
4	0.53-0.62	
5	0.58-0.67	Blue
6	0.61-0.72	
7	0.66-0.76	
8	0.70-0.93	Near IR
9	0.76-1.04	
10	0.90-1.38	
11	1.10-1.39	
12	1.30-1.79	
13	1.40-1.89	
14	1.90-2.39	
15	1.90-2.49	

6. Approach

All three classifiers were trained using the same training regions for each image. Providing an optimal, common training set for all classifiers was difficult but a quantitative comparison would not be possible without it. The accuracy of each of the resulting class maps was assessed using dependent, independent, and random reference sources. Reference data from DIRSIG material maps was used for the synthetic images as well. From these reference sources, the Simple, Weighted, Kappa, Prediger's Kappa, and the Tau coefficients were calculated. The results were obtained using a combination of real and synthetic imagery. The synthetic data sets served as a good indicator to bias in the other sampling techniques. Trends were then observed in the results obtained from both the sampling methods and accuracy metrics. The goal of this novel approach was to identify the optimal overall method for accuracy assessment of class maps based on accuracy and efficiency.

A single program was written to generate a confusion matrix and evaluate the five most common accuracy metrics. The confusion matrices were generated from any one of four different ground truth sources. Dependent, independent, random and synthetic data sets were read in as raw image files. In addition to any one of these data sets, the user must also supply a class map. This class map can be generated by any of the classification methods but must also be supplied in the form of a raw image file. Each reference and class map must be a single band image. Each class was designated by a unique digital count (DC) and the background class, if any, was designated by a DC of zero (black). The DC in the class map must match the DC in the truth data set for each corresponding class. This was done using a UNIX utility (XV) by changing the gray level in either image to match for each class. A key file was used for each class map to identify a class name with each DC.

6.1 Experimental Data Set Matrix

Three scenes were used as the basis for this effort. Images were generated from these scenes with degraded atmospheric visibility or spatial resolution. Images had three possible spatial resolutions: 1 meter, 2 meter, and 4 ground spot size. The atmospheric visibility of the images was either 23 kilometers, 7 kilometers, or 5 kilometers. Due to the large number of possible combinations of scenes and stressing parameters, only a limited number where selected for analysis. Figure 6-1 illustrates the experimental matrices for the stressing parameters of resolution and atmospheric visibility which were selected for each of the scenes. The figure indicates the source of the reference data used to assess the accuracy of each class map. The number next to the reference source indicates which classifier or classifiers was used to categorize that image. For each of the numbers, the scene was degraded, the classifier(s) were trained, the image classified, and the final class map accuracy was evaluated. As part of this thesis, a total of one hundred and nineteen (119) confusion matrices were generated. The results of these accuracy assessments, including confusion matrices, are contained in the appendices.

Tank Scene

Atmosphere \ Resolution	1m	2m	4m
23km	dependent 1,2,3 independent 1,2,3 scaled independent 1,2,3 random 1,2,3		
7km	dependent 1 independent 1 scaled independent 1 random 1		
5km	dependent 1 independent 1 scaled independent 1 random 1		

Forest Scene

Atmosphere \ Resolution	1m	2m	4m
23km	dependent 1,2,3 independent 1,2,3 scaled independent 1,2,3 random 1,2,3 synthetic 1,2,3		
7km	dependent 1,2,3 independent 1,2,3 scaled independent 1,2,3 random 1,2,3 synthetic 1,2,3		
5km	dependent 1,2,3 independent 1,2,3 scaled independent 1,2,3 random 1,2,3 synthetic 1,2,3		

Desert Scene

Atmosphere \ Resolution	1m	2m	4m
23km	dependent 1,2,3 independent 1,2,3 scaled independent 1,2,3 random 1,2,3	dependent 1,2,3 independent 1,2,3 scaled independent 1,2,3 random 1,2,3	dependent 1,2,3 independent 1,2,3 scaled independent 1,2,3 random 1,2,3
7km			
5km			

1 - Gaussian Maximum Likelihood
2 - Rule Based Genetic Algorithm (MYSTIC)
3 - Fuzzy ARTMAP Neural Network

Figure 6-1 Experimental Matrices

6.2 Importing Training Data

Training data consists of the digital counts (DC) in each band of a select pixel and the proper class to which it should be assigned. Training regions are the image areas over which training data is collected. Environment for Visualizing Information (ENVI™) was selected as the application from which training regions will be selected. Using the mouse, polygon vertices will be selected in each image to designate the desired classes. Different color polygons will be used for each class. These regions of interest (ROI) can then be used directly for supervised classification (GML) within ENVI™. To make an impartial comparison between classification algorithms it was decided common training sets would be used for each image with all three classification methods.

The following procedure was used to import training data into the Mystic™ rule base genetic algorithm classifier. First, the image underneath the polygons will be replaced by a black background within ENVI™. The ROI's superimposed over the black background will then be saved as a GIF image. This GIF image will then be converted to a portable pixel map (PPM) using PNMTOOLS. Once this is complete, the PPM image can be imported into a program which uses this image as a mask against the original image. Areas in the mask which are black are kept black, and in areas where the mask is not black the original image will pass. This will be done on each of the bands in the original image automatically by the program. The result is a Mystic™ training image which was black everywhere except were the desired ROI's were selected in ENVI™. In these areas, the original multiband image will appeared. Due to the Mystic™ 256 x 256 pixel limit on training images sizes, one extra step is required. The Mystic™ training images were larger than this so ENVI™ will be used to generate a smaller image (<256x256) into which each of the training regions will be cut and pasted. The mosaic image will serve as the final Mystic™ training image. This image is then imported into Mystic™'s training function. Each class region is selected by specifying the proper region from the Mystic™ training image. The Mystic™ iso-data function helps automate this processes by automatically selecting the proper polygon after the user clicks within each training region with the mouse. The iso-data parameters will be adjusted to the proper threshold to allow proper functioning. After the region selection is done, Mystic™ is trained (rules are optimized) on this data. Once Mystic™ has been trained, classification is performed on the original image. This procedure will be repeated for all the baseline images.

The procedure for training the Fuzzy ARTMAP will required less steps on the part of the user when compared to that of Mystic™. First the training polygons were saved as an image with a black background. Each of the polygons corresponding to an individual class will be designated by a unique color. This image is then saved from within ENVI™ as a GIF image exactly in the same manner as was used while training Mystic™. This image will then be converted into a PNM image where it is read into a program. This program also reads in the original image in the form of a band sequential (BSQ) raw file and outputs a training file in a format appropriate for the Fuzzy ARTMAP running under AVS.

6.3 Use of Synthetic Image Data

This thesis will utilize synthetic imagery to allow quantitative analysis of validation procedures and accuracy metrics. Synthetic Image Generation (SIG) allows researchers to simulate the image produced by a specific imaging system under various conditions on a computer without the time and expense involved in using the actual imaging system. The source of the synthetic images used in this thesis will be the Digital Imaging and Remote Sensing Image Generation (DIRSIG) package developed at the Rochester Institute of Technology (RIT) by the Digital Imaging and Remote Sensing (DIRS) Laboratory. DIRSIG utilizes a radiometrically accurate ray tracer to generate synthetic images from the visible to the long wave infrared regions of the electromagnetic spectrum. MODTRAN is used by DIRSIG to simulate atmospheric effects based on several meteorological conditions specified by the user. MODTRAN is an atmospheric radiation propagation model developed by the U.S. Air Force. DIRSIG has evolved over the years to include many additional features, which include a thermal sub-model, a sensor sub-model, and the ability to simulate MTF effects and object texture. For more complete coverage of DIRSIG features and theory, the reader is referred to *DIRSIG, Digital Imaging and Remote Sensing Image Generation, Description, Enhancement, and Validation* (Schott *et al*, 1993). Validation studies have indicated that images generated by DIRSIG can provide an accurate simulation of images taken with real sensors in the visible bandpass (White, 1996). While synthetic imagery will be employed for testing and comparing assessment techniques, it will not solely be used to determine the final characteristics and acceptability of these methods. For this purpose, highly characterized imagery, taken with real sensors will also be utilized. Two assumptions will be made when employing synthetic imagery for the purpose of evaluating classification accuracy metrics. First, if a metric yields poor results on synthetic images then it will perform at least as poorly on real images. The second assumption is good functionality on synthetic data does not ensure success with real images. This is why real images are being used as well. The synthetic data set will be used judiciously as a tool, but not an end all measure, to clarify and answer questions regarding the validity of computer generated images for this purpose.

Synthetic imagery has several properties which make it ideal for analyzing accuracy assessment techniques. It also does not suffer from many of the traditional disadvantages characteristic to other reference data collection techniques such as photointerpretation. Synthetic images generated by DIRSIG include a material map which details the exact material composition of each image pixel. Using the material map as the reference data, a population confusion matrix can be made for an entire classified image. This method of building a confusion matrix has several advantages over traditional approaches. Normally, sample confusion matrices are constructed from only a partial sample of the total population of image pixels and serve only as an estimator of total image statistics. With synthetic imagery, every pixel in the image is accounted for in the confusion matrix and there is no uncertainty in the accuracy of the reference data. With synthetic imagery the classified data is perfectly registered with the reference data, there is no temporal shift, and the minimum mapping unit (MMU) is exactly the same size as a class map pixel. Synthetic material maps will be used to generate sample and population size reference sets. Samples reference sets will be used to simulate typical reference sets gathered by a user.

6.4 Simulation of Stressing Parameters

The *desert* scene was used to analyze the effect of image resolution on classification accuracy and the *tank* and *forest* scenes were used to measure the effect of atmospheric attenuation. The stressing parameters were simulated in a different manner for all three scenes.

6.4.1 Modulation Transfer Function

Modulation Transfer Function (MTF) of an imaging system describes how it can discern fine spatial detail. This function will determine the effective ground instantaneous field of view (GIFOV) of a remotely sensed image. The GIFOV refers to the size of a single image pixel projected back onto the ground. The GIFOV determines the resolution of an image. The *desert* scene was selected to study the effect of resolution because of its large size. The original image had a GIFOV of approximately 1 meter. This image was convolved with a 2x2, equally weighted, convolution kernel to simulate the 2 meter GIFOV image. The 4 meter GIFOV image was generated in the same manner with a 4x4 kernel. Both convolved images were resampled using nearest neighbor interpolation to ½ and ¼ of the original image dimensions respectively. Degrading the image using equally sized convolution kernels on each spectral channel assumes the original image resolution was detector size rather than diffraction limited.

6.4.2 Atmospheric Effects

The atmospheric effects stressing parameter was simulated using two different approaches. The first approach was taken using the synthetic *forest* scene. DIRSIG, the

SIG package used to generate the *forest* scene, uses MODTRAN as its atmospheric propagation submodel. The three *forest* images were created by adjusting the MODTRAN inputs. These atmospheres were created based on inputs representing June 23, 1992 12:00 GMT at Rochester, New York. Atmospheres were simulated using a 23 kilometer visibility parameter (clear), 7 kilometer visibility (moderately clear), and 5 kilometer visibility parameter with 50% increase of relative humidity (hazy).

A second approach of simulating atmospheric visibility was employed with the *tank* scene. These atmospheres were simulated using a post-processing method. A linear histogram operation was used which reduced the scene contrast. To maintain consistency with the DIRSIG simulated imagery, the same MODTRAN inputs to the DIRSIG simulated atmospheres were used to modify the original image. The specific steps involved in converting the original imagery to atmosphere simulated imagery required the conversion of the original digital counts back to scene reflectance and applying the radiance reaching the sensor equation given by Equation 6-1.

$$L_\lambda = E_s' \tau_1(\lambda)\cos(\sigma)\frac{r(\lambda)}{\pi}\tau_2(\lambda) + \tau_2(\lambda)\, r(\lambda)\, L_{d\lambda} + L_{u\lambda} \qquad (6\text{-}1)$$

Where: E_s' is the exo-atmospheric irradiance $[\mathrm{w / m^2 \mu m}]$,
$\tau_1(\lambda)$ is the spectral transmission along the sun-target path,
$\tau_2(\lambda)$ is the spectral transmission along the target-sensor path,
σ is the solar declination angle,
$r(\lambda)$ is the spectral reflectance of the object,
$L_{d\lambda}$ is the downwelled solar radiance (skylight) $[\mathrm{w / m^2\, sr\, \mu m}]$ and,
$L_{u\lambda}$ is the path radiance $[\mathrm{w / m^2\, sr\, \mu m}]$.

This was followed by a conversion from radiance to digital counts based on M7 sensor gains and offsets for the various bands. Because these sensor gains and offsets were not available, an assumption was made to equate the original image with the clear atmosphere case. Using the set of gray panels of known reflectance within the *tank* scene, a radiance reaching the sensor was computed for the clear atmosphere case. Since the digital counts are also known (based on the clear atmosphere assumption) and assuming a linear relationship between radiance and digital counts, a nominal sensor gain and offset can be computed for each individual band. These gains and offsets were then used to consistently convert the atmosphere modified radiance to digital counts.

7. Theory

7.1 Factors Effecting Classification Accuracy

Many non-imaging factors effect the true and/or measured classification accuracy. The accuracy of supervised classification depends largely on a skilled analyst. For example, an experienced user can identify classes whose spectral probability distributions are bimodal and can split them in to two unimodal classes which can be recombined after classification. The importance of user supplied high, quality training data on supervised classification accuracy cannot be understated. The number of classes selected by the user in which to categorize the image will also have a strong bearing on the ultimate accuracy.

In addition to user factors, errors are introduced into classification by limitations of certain classifiers. For instance, the GML classifier assumes the DC distributions of target classes are Gaussian. Therefore, pixel distributions which differ significantly from Gaussian can not be classified accurately. Mixed pixels are pixels which have a combination of constituent signatures and are a large source of classification error. Single pixels around the borders of classes or in highly heterogeneous areas of an image have spectral contribution from multiple materials. Traditional classifiers have no special allowance for handling pixels of this type and they are often misclassified.

The number of pixels which a classifier leaves unclassified will also have a bearing on the measured classification accuracy. In general, pixels in which a classifier has difficulty segmenting will have lower than average classification accuracy. For this reason most classifiers incorporate a certainty threshold which must be exceeded before any given pixel is segmented. Classifiers such as the GML and Fuzzy ARTMAP allow the user to adjust this threshold while other classifiers such as Mystic™ do not. Pixels which a classifier leaves undefined do not contribute to classification accuracy assessment. Under most circumstances for a given image, training set, and classifier, the more pixels which are left undefined, the higher the measured classification accuracy. For this reason, classifiers which are forced to classify all pixels will results in lower measured classification accuracy. During supervised classification a balance must be struck between an acceptable number of unclassified pixels and the introduction of higher misclassification rates.

The information available to traditional classifiers, such as the one used in this thesis, are limited compared to human interpreters. While human interprets can classify pixels using the eight photointerpretation keys (Avery and Berlin, 1985), their computer counterparts commonly use only one key, tone.

7.2 Assessing Classification Accuracy

Many different approaches to classification validation have been presented in remote sensing literature. The techniques are usually implemented with a particular application in mind for the data set (Janssen and Van der Wel, 1994). Almost always, accuracy assessment consists of two steps: generating a confusion matrix and calculating an accuracy metric from that matrix. There are many variations of both steps. Several methods for gathering data to generate confusion matrices are presented in §7.3 and several accuracy metrics are defined in §7.4.

In the context of accuracy assessment of class maps, an error of commission refers to the classification of a pixel to a class in which it does not belong. An error of omission is committed by failing to assign a pixel to a class in which it rightfully belongs. There are several other terms which are used which correspond to these same types of errors. These terms often relate to the field in which the class map will be used. For example the defense industry often uses the terms false alarms and misses when referring to errors of omission and commission respectively.

$$p(y = Y|x = X) \quad ; \quad X,Y \in \{A,B,C\} \tag{7-1}$$

Where: x is the true class,
y is the classification results and,
A, B, C are class categories.

Figure 7-1 is a contingency diagram illustrating the probabilities a classifier will make different types of errors. The pixels on the left are categorized into the classes **A**, **B**, and **C** on the right by the classifier. The variable **x** is the true class and **y** is the class in which each pixel is classified. For instance, the probability that a pixel which is really from class **A** will be classified as class **C** is p(y=C|x=A). This conditional probability approach will be used by some accuracy metrics.

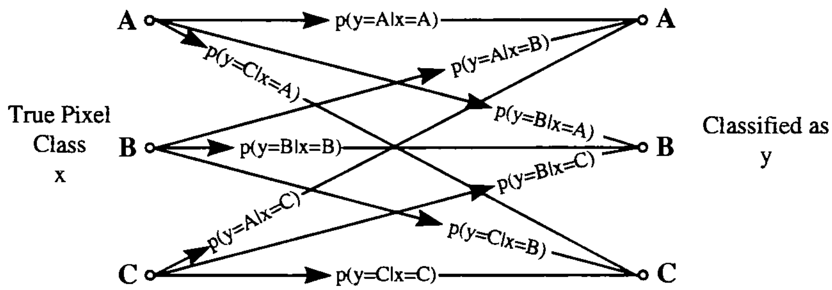

Figure 7-1 Contingency Diagram

A more useful probability than the chance a pixel from a known class is misclassified [p(y=Y|x=X)] is the chance that a classified pixel is classified correctly [p(x=X|y=Y)]. This probability can be found using Bayes theorem (Equation 7-2). Assuming proper sampling of a valid reference source, all the needed terms can be taken from a confusion matrix. Confusion matrices and the conditional probabilistic approach will be discussed in the next section.

$$p(x = X|y = Y) = \frac{p(y = Y|x = X)\, p(x = X)}{p(y = Y)} \tag{7-2}$$

Understanding the spatial orientation of errors could help determine the most efficient method for measuring classification error. Spatial autocorrelation analysis has indicated that class maps usually exhibit positive autocorrelation (Congalton, 1988). This means errors are clumped or clustered together rather than randomly distributed throughout the image. This fact has serious implications in that care must be taken in selecting an appropriate sampling method for evaluation. Several common sampling methods are discussed in §7.3.2 for this very reason. Accuracy assessment never results in a definite figure of merit but rather an estimate of the overall classification accuracy of which the confidence interval can be known. The exact accuracy cannot be know with perfect confidence because confusion matrix reference data is known only for a finite sample of the population of image pixels.

7.3 Confusion Matrices

A confusion matrix is a square table indicating the classes to which pixels where assigned and the classes to which they should have been assigned. Different authors have referred to them as error, contingency, evaluation, misclassification, and accuracy matrices or tables. There are two types of confusion matrices, the sample and the population confusion matrix. The far more common type, sample confusion matrices, are constructed from only a sampling of pixels from the class map being verified. Population confusion matrices are tables in which all the class map pixels are accounted for. In this study, population confusion matrices were generated using synthetic image information as a reference source. Accuracy metrics calculated using a sample confusion matrix are based on incomplete sampling and will have a corresponding uncertainty. Metrics calculated using a population matrix are exact measurements and do not therefore have a confidence interval. Population confusion matrices are virtually impossible to generate under normal circumstances.

Several sampling methods exist for gathering data which can be used to generate a sample confusion matrix. These methods are divided into six distinct types and presented in §7.3.1 and §7.3.2. Each pixel from the sample is then placed in the appropriate row and column of the matrix. Traditionally, reference data is placed in columns and classification results are place in the rows. This means the value in the matrix column corresponding to the class category of the reference data set and row corresponding to the class map category is incremented by one for each pixel location in the sample. The matrix is discrete and is always square with size M x M, where M is the number of class categories. The numbers along the main diagonal of the square matrix indicate correctly classified pixels while all off-diagonal elements are misclassified. The total number of pixel counts in the confusion matrix is almost always limited by the size of the reference sample. From this sample of pixels, the accuracy of the entire image can be approximated. Ideally, the sampling method will select representative pixels and the sample will serve as an unbiased estimator of the entire image. This is often not the case and results are often skewed for reasons which will be discussed later in this section. Without an accurate confusion matrix classification accuracy assessment is not possible.

Reference Data

Classified Data	256	Coniferous	Deciduous	Scrub	Concrete	Road	Sand	$i+$
	Coniferous	41	13	1	0	0	2	57
	Deciduous	8	92	1	0	4	0	105
	Scrub	1	3	18	0	0	1	23
	Concrete	4	0	0	16	1	4	25
	Road	1	2	0	4	5	2	14
	Sand	2	1	0	0	1	28	32
	$+i$	57	111	20	20	11	37	200

Figure 7-2 Sample Confusion Matrix

Categorical data analysis is used to analyze contingency tables such as confusion matrices. Several categorical data analysis methods, accuracy metrics in this case, will be presented in §7.4. Square categorical data of any type often has two noteworthy properties (Agresti, 1990). First, they often exhibit a symmetric pattern about the main diagonal. In the case of confusion matrices this is due to confusion which arises during classification from two class which are spectrally similar. During classification if a classifier frequently misclassifies a pixel of class **A** as class **B** there is also a high probability that the same classifier will frequently misclassify a pixel of class **B** as class **A**. This phenomena results is symmetry about the main diagonal. The second property is largely caused by this occurrence. The two marginal distributions frequently differ systematically. The marginals are simply the summation along the corresponding row or column.

Reference Data (X)

Classified Data (Y)	A	B	C	
A	$p(y=A \mid x=A)$	$p(y=A \mid x=B)$	$p(y=A \mid x=C)$	$p(y=A)$
B	$p(y=B \mid x=A)$	$p(y=B \mid x=B)$	$p(y=B \mid x=C)$	$p(y=B)$
C	$p(y=C \mid x=A)$	$p(y=C \mid x=B)$	$p(y=C \mid x=C)$	$p(y=C)$
	$p(x=A)$	$p(x=B)$	$p(x=C)$	

Figure 7-3 Probabilistic Confusion Matrix

A sample 6 x 6 confusion matrix is shown in Figure 7-2. The reference and classified categories are list along the top and left side respectively. It should be noted that researchers have several names for these categories. Classified data is sometime known as predicted, evaluated, interpreted, or observed data and reference data is sometime called verified, identified, known, or truth data. The values along the far right column and bottom row are the row and column marginals. The sum of all the values in the matrix is equal to the total number of pixels in the sample size, that is, all the pixels being evaluated. This value (256) is indicated in the upper left hand corner of this matrix. The summation of the values along the diagonal of this matrix (200) has been posted in the lower, right hand corner. This is the total number of elements properly classified.

It is often useful to use a probabilistic interpretation of a confusion matrix as shown in Figure 7-3. In this case, pixel counts are replaced by conditional probabilities of occurrence. This means the **p(y=A|x=B)** is the conditional probability that a pixel is classified as class **A** given the fact that the pixel truly belongs to class **B**. These probabilities are found by simply dividing the pixel counts in each category by the matrix wide sum (*N*). The marginals are replaced by the proportion of each class. For example, **p(y=A)** is the probability that a pixel in the sample being evaluated was classified as class **A**. All the conditional probabilities, in addition to the row and column probabilities, each sum to unity. Several sampling methods, for obtaining reference data, will be discussed below.

7.3.1 User Selected Reference Data

User selected reference is a type of reference which is collected by an image analyst. The analyst select not only the location of the regions to verify but also their size and shape. With this type of sampling, large polygons are usually selected over areas of the image which the analyst deems homogeneous. In most implementations, selection of user reference is accomplished within a graphical user interface (GUI) where the analyst selects the endpoints of polygon vertices. Two types of user selected reference are often encountered, dependent and independent data sets.

7.3.1.1 Dependent Data Sets

The first sampling method involves the use of a dependent data set. With this method the polygons of contiguous pixels, which were originally used to train the classifier, are used to evaluate the accuracy of the resulting class map. The confusion matrix is then gathered by determining whether the class of each pixel in the training polygons matches that of the class map. This method is the easiest and also the least trusted. This type of accuracy assessment can even be automated so a classifier returns its own classification accuracy, without supplying any additional information.

The results obtained will almost always overestimate the true accuracy of the classifier. There are several reasons for this. The training regions designated by a user are usually selected because they are some of the most homogeneous class areas in an image. Using the dependent data set assesses the accuracy only in these relatively homogeneous areas. More heterogeneous areas, where the classifier is more likely to err, are neglected altogether. This high correlation leads to the overly optimistic confusion matrix inherent to dependent data set evaluations. In addition, because the classifier is trained on the dependent data it has the highest accuracy in this region. Iterative classifiers, such as Mystic™, actually use dependent set accuracy as a feedback mechanism into the classifiers. The true information content conveyed by each pixel in dependent data is less than it would be otherwise because the majority of the pixels are contiguous. Contiguous pixels are highly correlated with respect to class membership and classifier error. Optimistic bias in reference data is further discussed in §8.1.

Even though dependent skewed reference data is generally accepted to be inaccurate, it is still used by many researchers (Franklin and Wilson, 1992; Bauer *et al*, 1994; Nageswara Rao and Mohankumar, 1994). It is still used because it is convenient and many smaller project cannot handle the cost or time requirements of more rigorous verification efforts.

7.3.1.2 Independent Data Sets

The second method is similar to the first but it uses an independent data set of polygons. With this technique the user designates a second set of polygons, independent of the training regions, which correspond to samples of each class. The confusion matrix is then generated in the same manner as the dependent set technique except the independent regions are used. This method is not as biased as the previous because the regions are not correlated with classifier training. However, correlation is still a problem due to the use of contiguous pixels. When selecting independent data sets, care should be taken on the part of the user to avoid image areas which were used to train the classifier. Independent data sets, like dependent sets, do not yield accurate *a priori* class probabilities. While this sampling technique has less bias, it is also more time consuming because a second set of polygons is needed.

7.3.2 Random Point Sampling

Confusion matrices are constructed by taking pixel samples from a population of image pixels. For reasons of practicality, the accuracy of all the pixels in an image cannot be verified, with the exception of synthetic imagery. If reference data is already known about every pixel in the image then classification would be pointless. Confusion matrices are usually generated by spot checking pixels with various methods using different sampling to select pixel locations. Many different techniques of spot checking have been utilized by the remote sensing community. Some studies have used a helicopter to physically visit predetermined sites and visually verify landcover while other efforts have employed an aircraft and used a video recorder to capture ground cover information (Szajgin *et al*, 1982; Marsh *et al*, 1994). By far, the most common technique for verifying ground cover is visual interpretation of aerial photographs. This widely accepted method utilizes a trained observer to categorize ground areas for use as reference data. Identification of objects in a scene is based on human recognition employing the nine interpretation keys: pattern, tone, texture, shadow, site, shape, size, association, and resolution (Avery and Berlin, 1985). Human interpretation is often aided by viewing stereoscopic image pairs which can provide further spatial cues. Human interpretation of aerial photographs has proven to be a highly accurate process.

However, before photointerpretation or physical verification can be done, the location of the samples must be predetermined. Many techniques have been employed to determine the location and number of pixels to be sampled. If any sampling technique selects a pixel which a classifier has left undefined then this pixel is not used and another pixel is selected in its place. Random verification is usually accomplished the quickest if sample locations are first sorted by location. This aids the analyst in quickly identifying familiar ground cover. There are several common sampling techniques such as simple random sampling, stratified random sampling, clustered sampling, systematic sampling, and stratified systematic unaligned sampling (Congalton, 1988).

7.3.2.1 Simple Random Sampling

Simple random sampling refers to selecting single pixels at random such that each pixel has an equal chance of being selected. This type of sampling can be used with or without replacement. If replacement is used a single pixel in the population image could be selected more than once.

7.3.2.2 Stratified Random Sampling

Stratified random sampling is a popular technique which ensures small classes, which might have been missed by truly random sampling, are represented in the confusion matrix. With this technique *a priori* knowledge about the image population is used to divide the image into non-overlapping subpopulations, or strata, for each class category. Random, single pixels, samples are then taken from each strata. This sampling method is popular due to several beneficial properties. First and foremost, all class categories can be sampled a sufficient number of times while maintaining a tractable number of total samples. A user might also only be interested in a single class. With stratified random sampling, accuracy assessment can be performed on that single class by only sampling that single strata.

7.3.2.3 Clustered Random Sampling

Clustered sampling is the same as simple random sampling except rather than a single pixel, a group of pixels in one area is selected. While this technique is statistically inferior to completely random sampling it is commonly employed due to logistical limitations. With simple random and clustered sampling the same amount of samples can be generated while visiting far fewer sites with the latter.

The first group to employ clustered sampling, for the purpose of assessing digital classification accuracy, was the U.S. Geological Survey's EROS Data Center in 1976 (Linden and Szajgin). This approach is considered an efficient means of determining the constituent class of image pixels when undertaking physical identification. The majority of expense incurred in any type of physical identification takes place while traveling to and locating the regions of interest. The cost of data acquisition, once at the appropriate site, is trivial in comparison. For this reason, with clustered sampling, after traveling to the region of interest, several samples in close proximity are taken. Several samples are then taken in aggregate to form the minimum mapping unit. This technique then avoids the marginal expense of additional transit to another remote site, in effect reducing the per sample cost. The advantage of cost savings with this method of sampling is moderated by a major disadvantage associated with it. From a statistical standpoint, additional samples are advantageous because they decrease the associated confidence interval of an approximation. However, while this technique can obtain more samples for a fixed cost, the samples are not completely random as our statistics assume. This problem is further aggravated by an observation noted in §7.2 and later discussed in §8.1. Errors in class maps, as indicated by Congalton (1988), often exhibit strong positive spatial correlation. Clustered samples are therefore not independent random samples, an assumption of many accuracy metrics, due to the high correlation between samples in close proximity.

7.3.2.4 Systematic Random Sampling

Systematic sampling is a method whereby single pixel samples are taken from locations at equally spaced intervals over the entire image. First the location of the starting point, or first sample, is selected at random and the spatial interval is specified by the user. Each successive sample is taken at a fixed interval thereafter. Systematic sampling is used to ensure that a spatially uniform spread of samples are taken over an entire image and to preclude the chance that a spatial isolated area is overly represented in the measured accuracy metric. Systematic sampling is statistically flawed in that the selection procedure implies that each unit in the image population has an equal chance of being included in the sample (Congalton, 1988). Large systematic biases can be encountered with this technique if the image contains any periodicity.

When validating classification results the minimum mapping unit (MMU) is often larger than a pixel size in the classified image. MMU refers to the smallest area visually interpreted on a reference image. In some studies, the MMU is selected to have dimensions of some significance inherit to the application at hand. For example, the U.S. Department of Agriculture Forest Service has used MMU's defined to be the size of individual illuminated tree crowns when assessing the accuracy of tree species classification (Thomasson *et al*, 1994). The principal sampling unit (PSU) is similar to the MMU. The PSU refers to the smallest area used to train a classifier. In some cases, where a dependent data set is utilized as reference data, the PSU is the same as the MMU.

The binomial probability density function can be used to model the expected number of misclassification occurrences (omission and commissions) from a given sample size. With a class map of accuracy p there will be a probability f of observing y misclassifications given a n pixel sampling size.

$$f[y] = \frac{n!}{(n-y)!\,y!} p^{y} p^{n-y} \qquad (7\text{-}3)$$

Where: n = sample size,
y = number of misclassifications, and
p = true map accuracy.

Using the binomial distribution to approximate sampling is statistically valid only when independent random samples are taken. This technique is not appropriate when samples are not taken at random such as the dependent and independent data sets described in §7.3.1.1 and §7.3.1.2. This type of sampling introduces bias and systematic error. Other researchers have instead employed the normal distribution as the large sample approximation to the binomial (Rosenfield and Melley, 1980). These distributions can be used to determine the number of samples required to achieve a desired confidence interval. The binomial approximation is not necessary when using synthetic imagery because reference data is available for the entire image population, not just a finite sample.

7.3.3 Synthetic Imagery Verification

In this study synthetic imagery was used to evaluate the performance of the sampling techniques and accuracy metrics. Verification was done in two ways, from a small sampling and with the entire image population. The sampling techniques will be used with synthetic material map as the reference source. A sample confusion matrix will then be generated from this data in the normal fashion. A population confusion matrix will also be generated from all the pixels in the image using the material map as the reference source. The error metrics calculated from these matrices, found in the appendix, are exact measurements and have no confidence interval. In Figure 7-4 the material map is shown on the far left, the *forest* scene in the middle, and the regions used to train the classifiers on the far right.

Figure 7-4 Classification and Verification of Synthetic Scene

DIRSIG provides a series of debugging images with each synthetically generated image. These images are normally used for solving problems or better understanding the resulting synthetic image. In this project the material map debug image was used to determine the exact composition of each pixels. Converting the DIRSIG material map to accuracy assessment reference data required two main steps. First the DIRSIG materials were mapped to the categories that had been chosen for supervised classification. The exact conversion is shown in Figure 7-5.

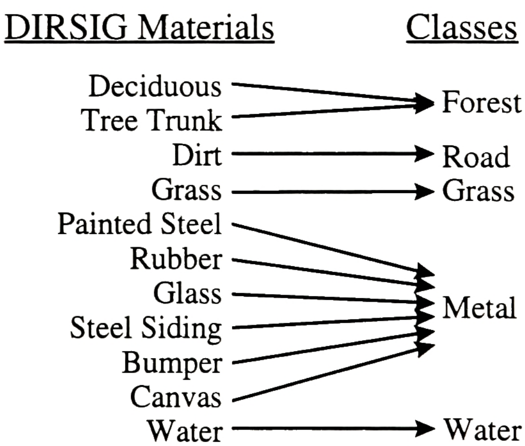

Figure 7-5 Mapping of DIRSIG Materials to Class Map Categories

The second step was an operation to compensate for the convolution and down-sampling that were used to covert the DIRSIG radiance field into the final *forest* scene. The mode, the observed material with the highest observed frequency, was taken from a 3x3 window run across the entire material map. The result of this operation was posted in a new image one third the size of the material map. This new image was the reference map. Next, the window was moved 3 pixels and the process was repeated. When ties of the mode operation were encountered, the window was expanded to 4x4 and the mode of this window was used. The window was expanded until there was no tie. Due to the size of the window, the edges around the reference map were cut off, the same as the *forest* image.

7.4 Accuracy Metrics

Ideally, classifier performance or class map accuracy could be summarized with a single, intuitive metric. This ideal metric would not over or under estimate the true accuracy and it would account for all types of error. Unfortunately, the ideal metric does not exist. Several metrics will be presented in this section which use only a confusion matrix and categorical data analysis to estimate classification accuracy. Errors in the confusion matrix will lead to poor results from any metric. There are two distinct types of accuracy metrics. The first, and most common type, are overall image metrics. They result in a single number which estimates the accuracy of the entire image and all classes. The second type are the single class metrics. These metrics provide estimates of the accuracy for each class. All accuracy metrics are estimators of a true value. As such, the have a degree of uncertainty which is usually expressed as an interval with a specified level of confidence.

7.4.1 Uncertainty of Estimates and Confidence Intervals

All classification accuracy metrics provide a single number estimate as an indication of their quantitative assessment whether it is for a specific class or the entire image. Each of these metrics has a corresponding variance associated with it. Reporting the appropriate confidence interval is therefore an essential part of reporting results of this type. The confidence interval is a range of the metric values which has a fixed probability of containing the true value. As a side note, the confidence intervals reported in remote sensing literature is almost exclusively due only to incomplete sampling of a population. There are, in fact, many other sources of error which are typically not accounted for such as human error in photointerpretation and biased sampling.

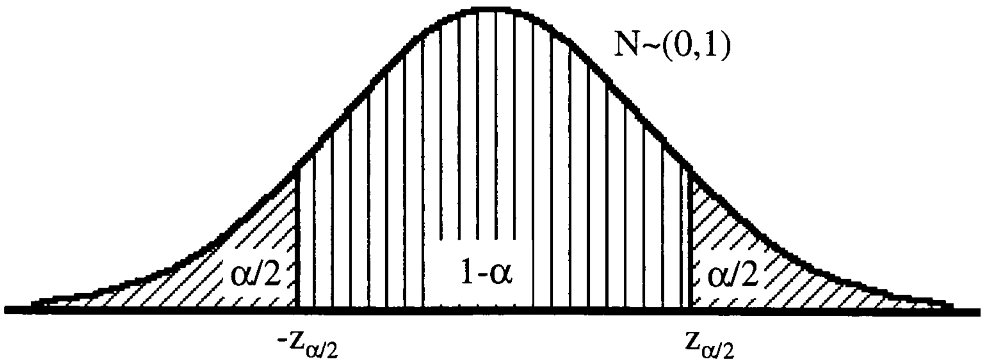

Figure 7-6 Standard Normal Density

The first step is finding the confidence interval is deciding on the desired confidence. A typical value is 95% which will be used for the examples for each metric. This probability is denoted by $1-\alpha$. The corresponding z-score ($z_{\alpha/2}$) can then be found from a table of the standard normal density. Table 7-1 summarizes several commonly used confidence intervals.

Table 7-1 Confidence Interval Z-Scores

$1-\alpha$	0.80	0.85	0.90	0.95	0.99
$z_{\alpha/2}$	1.28	1.44	1.64	1.96	2.58

Finally, the confidence interval can be found by using Equation 7-4.

$$\hat{x} \pm z_{\alpha/2}\sigma \tag{7-4}$$

Where: $\hat{x}$ = metric value,
$z_{\alpha/2}$ = z-score , and
σ = metric standard deviation.

This formula assumes a large sampling of at least 200 points.

7.4.2 Image Wide Accuracy Metrics

Below are all of the popular accuracy assessment metrics. Each of the metrics will be described along with any underlying assumptions. The formula of the metric with the appropriate variance will be given as well. Finally a numerical example, with the confidence interval, will be calculated for each metric from the sample confusion shown in Figure 7-2.

7.4.2.1 Simple Accuracy

The simplest and most easily understood metric for indicating classification accuracy is the Simple accuracy metric (P_o) given in Equation 7-6. Often known as Proportion of pixels Correctly Classified (PCC) (Veregin, 1989) it is found by taking the ratio of correctly classified pixels to the total number of pixels in the confusion matrix. The number of correctly classified pixels is found by taking the summation of all the pixels along the diagonal from the upper-left to lower-right corners of the confusion matrix. The metric is equally as easy to interpret. It is the probability that any pixel in the sampled set, selected at random, will be properly classified. As a true probability it has a minimum value of 0.0 and a maximum value of 1.0. The total number of pixels in a confusion matrix (N) is found by:

$$N = \sum_{i=1}^{M} \sum_{j=1}^{M} x_{ij} \tag{7-5}$$

Where: M is the number of classes, and
x_{ij} is number of pixels falling into both categories of the i^{th} row and j^{th} column.

$$P_o = \frac{1}{N} \sum_{i=1}^{M} x_{ii} \tag{7-6}$$

$$N \sim (NP_o, \sqrt{NP_o(1-P_o)}) \tag{7-7}$$

The variance of percent correct metric (P_o) is found by:

$$\sigma^2[P_o] = \frac{1}{N} P_o(1-P_o) \tag{7-8}$$

While the percent agreement metric is a simple and intuitive measure of agreement, it does have an important disadvantage. The percent agreement metric overestimates thematic map accuracy because it does not account for the proportion of agreement between the classified and reference data sets that is due to chance agreement alone (Congalton and Mead, 1983; Congalton *et al.*, 1983). The other metrics presented in this section attempt to correct for this shortcoming but use the Simple accuracy as a basis.

A sample calculation is shown below for the confusion matrix Figure 7-2. The percent agreement metric is evaluated and the appropriate variance is found. This information is also used to find the resulting 95% confidence interval.

The percent agreement metric using Equation 7-6:

$$P_o = \frac{1}{256}(41+92+18+16+5+28) = 0.781 = 78.1\%$$

The percent agreement variance using Equation 7-8:

$$\sigma^2[P_o] = \frac{1}{256}0.78125\,(1-0.78125) = 6.68\text{x}10^{-4}$$

The 95% confidence interval of $P_o = 0.781 \pm (1.96)\sqrt{6.68\text{x}10^{-4}} = (0.731, 0.832)$.

7.4.2.2 Weighted Accuracy

Weighted Accuracy (Equation 7-9) is similar to the Simple accuracy but each class category is weighted equally. The result is the average percent accuracy of all the classes assuming all classes are considered equally important. This value can be larger or smaller than that of the Simple accuracy. The Weighted accuracy metric has not been as widely used as other metrics in the remote sensing literature. Under certain circumstances the Weighted accuracy could provide more accurate results than the simple percentage. The Simple accuracy metric assumes that the proportion of reference pixels (column marginals) in a certain class is representative of the true proportion in the image. Using stratified random or user selected sampling methods this may not be the case. If a class if poorly classified, and over represented in the reference set, the Simple accuracy metric will underestimate the true accuracy. The Weighted accuracy, in cases where the *a priori* class probabilities are not known, may obtain better results.

$$w = \frac{1}{M}\sum_{i=1}^{M}\frac{x_{ii}}{x_{+i}} \qquad (7\text{-}9)$$

Where: '$_+$ 'represents a summation over that index.

The confidence interval for the Weighted accuracy metric can be derived by first letting X_i be equal to the true number of properly classified pixels in a sample of n pixels from class i. Then we can treat X_i a binomial variable (Equation 7-10) with parameters n_i and P_i.

$$X_i \sim \text{bino}(n_i, P_i) \tag{7-10}$$

Where: n_i is the number of pixels in class i, and
P_i is the proportion of properly classified pixels.

The true portion of properly classified pixels in each class (P_i) can be estimated using the Producer's accuracy ($\hat{P}_i$). The Producer's accuracy will be discussed in greater detail later in this section but is for now defined using Equation 7-11.

$$\hat{P}_i = \frac{x_{ii}}{x_{+i}} \tag{7-11}$$

Where: x_{ii} is number of pixels in the i^{th} row and i^{th} column, and
x_{+i} the column marginal.

The column marginals (x_{+i}) are simply the column sums and can be found using Equation 7-12.

$$x_{+i} = \sum_{j=1}^{M} x_{ji} \tag{7-12}$$

Next, we may find the distribution of the Producer's accuracy by the DeMoivre-Laplace limit theorem approximation of the binomial distribution by the normal distribution (Equation 7-13).

$$\hat{P}_i = \frac{X_i}{n_i} \sim N\left(P_i, \sqrt{\frac{P_i(1-P_i)}{n_i}}\right) \tag{7-13}$$

This result indicates that the Producer's accuracy will have a standard deviation given by: $\sqrt{\frac{P_i(1-P_i)}{n_i}}$.

Assuming that each of the Producer's accuracies are independent, the distribution of their sum is found by the summation of the normal distribution parameters (Equation 7-14).

$$\sum_{i=1}^{M} \hat{P}_i \sim N\left(\sum_{i=1}^{M} P_i, \sqrt{\sum_{i=1}^{M} \frac{P_i(1-P_i)}{x_{+i}}}\right) \tag{7-14}$$

Finally, the distribution of the Weighted accuracy metric is found (Equation 7-15).

$$w = \frac{1}{M}\sum_{i=1}^{M} \hat{P}_i \sim N\left(\frac{1}{M}\sum_{i=1}^{M} P_i, \sqrt{\frac{1}{M^2}\sum_{i=1}^{M} \frac{P_i(1-P_i)}{x_{+i}}}\right) \tag{7-15}$$

The standard deviation of the Weighted accuracy is therefore: $\sqrt{\frac{1}{M^2}\sum_{i=1}^{M} \frac{P_i(1-P_i)}{x_{+i}}}$.

$$P_i \approx \frac{x_{ii}}{x_{+i}} \tag{7-16}$$

We again rely on the fact that the true proportion of properly classified pixels in each class is closely approximated by the ratio of the diagonal element to the column sum of the corresponding confusion matrix column (Equation 7-16). Substituting this relation into the standard deviation equation yields the final variance of the Weighted metric (Equation 7-17).

$$\sigma^2[w] = \frac{1}{M}\sum_{i=1}^{M} \frac{x_{ii}}{x_{+i}^2}\left(1 - \frac{x_{ii}}{x_{+i}}\right) \tag{7-17}$$

A sample Weighted accuracy calculation is shown below for the confusion matrix in Figure 7-2.

$$w = \frac{1}{6}\left(\frac{41}{57}+\frac{92}{111}+\frac{18}{20}+\frac{16}{20}+\frac{5}{11}+\frac{28}{37}\right) = 0.743 = 74.3\%$$

$$\sigma^2[w] = \frac{1}{6}\left[\frac{41}{57^2}\left(1-\frac{41}{57}\right)+\frac{92}{111^2}\left(1-\frac{92}{111}\right)+\frac{18}{20^2}\left(1-\frac{18}{20}\right)+\frac{16}{20^2}\left(1-\frac{16}{20}\right)+\frac{5}{11^2}\left(1-\frac{5}{11}\right)+\frac{28}{37^2}\left(1-\frac{28}{37}\right)\right]$$

The 95% confidence interval of $P_o = 0.743 \pm (1.96)\sqrt{6.68\text{x}10^{-4}} = (0.731, 0.832)$.

7.4.2.3 Kappa Coefficient

Kappa Coefficient ($\hat{k}$) was first derived by Cohen (1960) as a measure of the agreement between categorical data from psychological testing. The Kappa coefficient was later introduced to the remote sensing community as a statistical measure of classification accuracy by Congalton *et al* (1980). In recent times the Kappa statistic has received a lot of attention and become the *de facto* standard metric for the approximation of classification accuracy assessment. It is a categorical data analysis index of agreement between two data sets when chance agreement is a concern.

In the past several published papers have made errors in the formulation of the Kappa statistic and the large sample variance (Hudson and Ramm, 1987). Several numerical errors have been published as well to add to the confusion. The correct formula for the Kappa statistic is given in Equations 7-18 and 7-19. These Equations are the two most commonly presented forms of the statistic and are algebraically equivalent. The former is usually found in mathematical publications while the latter is normally used for numerical calculations.

Unlike the percent agreement metric, the Kappa coefficient attempts to account for chance agreement by incorporating all marginal distributions of the confusion matrix (Cohen, 1960). However, Foody (1992) has shown that without modification Kappa overestimates this proportion of agreement due only to chance (P_c) and therefore underestimates the overall classification.

$$\hat{k} = \frac{P_o - P_c}{1 - P_c} \quad (7\text{-}18)$$

$$\hat{k} = \frac{N\sum_{i=1}^{M} x_{ii} - \sum_{i=1}^{M} x_{i+}x_{+i}}{N^2 - \sum_{i=1}^{M} x_{i+}x_{+i}} \qquad (7\text{-}19)$$

Where: N is the number of pixels in the matrix,
M is the number of classes,
x_{i+} the row marginal,
x_{+i} the column marginal, and
x_{ij} is number of pixels falling into both categories of the i^{th} row and j^{th} column.

The 'chance' term (P_c) is the proportion of pixels which agree due to chance.

$$P_c = \frac{1}{N^2}\sum_{i=1}^{M} x_{i+}x_{+i} \qquad (7\text{-}20)$$

The row marginals (x_{i+}) can be found using Equation 7-21.

$$x_{i+} = \sum_{j=1}^{M} x_{ij} \qquad (7\text{-}21)$$

The approximate large sample variance of the Kappa statistic is given by Equation 7-22. Two separate variables (Equation 7-23 and 7-24) are used to simplify the equation.

$$\hat{\sigma}^2[\hat{k}] = \frac{1}{N}\left[\frac{P_o(1-P_o)}{(1-P_c)^2} + \frac{2(1-P_o)(2P_oP_c - \alpha_1)}{(1-P_c)^3} + \frac{(1-P_0)^2(\alpha_2 - 4P_c^2)}{(1-P_c)^4}\right] \qquad (7\text{-}22)$$

P_o was defined in Equation 7-6 to be the proportion of pixels laying along the matrix diagonal. That is, the percent of pixels properly classified. With perfect agreement this value would equal 1.0.

$$\alpha_1 = \frac{1}{N^2}\sum_{i=1}^{M} x_{ii}(x_{i+} + x_{+i}) \qquad (7\text{-}23)$$

$$\alpha_2 = \frac{1}{N^3}\sum_{i=1}^{M}\sum_{j=1}^{M} x_{ji}(x_{i+} + x_{+j})^2 \qquad (7\text{-}24)$$

A sample Kappa coefficient calculation, using the first Equation (Equation 7-19), is shown below for the sample confusion matrix in Figure 7-2.

$$\hat{k} = \frac{256(41+92+18+16+5+28)-(57\cdot57+105\cdot111+23\cdot20+25\cdot20+14\cdot11+32\cdot37)}{256^2-(57\cdot57+105\cdot111+23\cdot20+25\cdot20+14\cdot11+32\cdot37)} = 0.703 = 70.3\%$$

$$P_c = \frac{1}{256^2}(57\cdot57+105\cdot111+23\cdot20+25\cdot20+14\cdot11+32\cdot37) = 0.2625$$

$$\alpha_1 = \frac{1}{256^2}\left[41\cdot(57+57)+92\cdot(105+111)+18\cdot(23+20)+16\cdot(25+20)+5\cdot(14+11)+28\cdot(32+37)\right] = 0.4287$$

$$\alpha_2 = \frac{1}{256^3}(810569+4727475+57638+37584+69382+166368) = 0.350$$

$$\hat{\sigma}^2[\hat{k}] = \frac{1}{256}\left(\frac{0.7812\cdot(1-0.7812)}{(1-0.2625)^2}+\frac{2\cdot(1-0.7812)\cdot(2\cdot0.7812\cdot0.2625-0.4287)}{(1-0.2625)^3}+\frac{(1-0.7812)^2\cdot(0.35-4\cdot0.2625^2)}{(1-0.2625)^4}\right)$$
$$= 1.20\text{x}10^{-3}$$

The 95% confidence interval of $\hat{k} = 0.703 \pm (1.96)\sqrt{1.20\text{x}10^{-3}} = (0.635, 0.771)$.

7.4.2.4 Brennan and Prediger's Kappa

Brennan and Prediger's Kappa coefficient ($\hat{k}_n$) was introduced by Brennan and Prediger (1981) for measuring the agreement between two data sets which accounts for a process but neglects chance agreement. This categorical data analysis metric was originally used for studies of reliability and validity. This modified Kappa statistic was introduced to the analysis of confusion matrices by Foody (1992). The motivation behind this new metric is the belief that the standard Kappa coefficient underestimates true classification accuracy by over inflating the chance agreement term (P_c).

The Kappa chance term (Equation 7-20) is appropriate when the marginals are 'fixed' (Brennan and Prediger, 1981). Fixed marginals refers to circumstances were marginal proportions are known to the classifier before classifying the pixels to classes. In cases where the marginals are 'free', the Kappa chance term should be replaced with the reciprocal number of classes (Brennan and Prediger, 1981). Marginals are considered free when they are not known *a priori* to the classifier, as is the case for most classifiers. For this reason the revised Kappa statistic is give in Equation 7-25.

$$\hat{k}_n = \frac{P_o - \frac{1}{M}}{1 - \frac{1}{M}} \qquad (7\text{-}25)$$

Where: M is the number of class categories.

The Kappa magnitude of the Kappa chance term (P_c) includes both chance agreement and actual classifier agreement. However, the new chance term, reciprocal number of classes, includes only chance agreement. With this new term the Brennan and Prediger's Kappa coefficient ($\hat{k}_n$) is a measure of the portion of properly classified pixels due only to classifier performance assuming the marginals are free. As such, the modified Kappa statistic usually results is a larger percentage of agreement than does the standard Kappa coefficient.

The variance of the Brennan and Prediger's Kappa coefficient ($\hat{k}_n$) can be found using Equation 7-26. It is the same as the Tau variance (Equation 7-32) but the random agreement (P_r) term is replaced by the reciprocal number of class categories.

$$\sigma^2[\hat{k}_n] = \frac{P_o(1 - P_o)}{N(1 - \frac{1}{M})^2} \qquad (7\text{-}26)$$

Below is a sample Brennan and Prediger's Kappa coefficient calculation with the appropriate 95% confidence interval for the sample confusion matrix data in Figure 7-2.

$$\hat{k}_n = \frac{0.78125 - \frac{1}{6}}{1 - \frac{1}{6}} = 0.738 = 73.8\%$$

$$\sigma^2[\hat{k}_n] = \frac{0.78125(1-0.78125)}{256(1-\frac{1}{6})^2} = 9.61\text{x}10^{-4}$$

The 95% confidence interval of $\hat{k}_n = 0.738 \pm (1.96)\sqrt{9.61\text{x}10^{-4}} = (0.677, 0.798)$.

7.4.2.5 Tau Coefficient

The Tau coefficient (T), first introduced by Klecka (1980), measures the improvement in chance agreement between two data sets over random assignment. Recently, the Tau statistic has been applied to classification accuracy assessment by Ma and Redmond (1995). It is their contention that this metric can address some of the disadvantages of other statistical metrics, particularly the Kappa coefficient. The proposed advantages over Kappa include better adjustment of percent agreement, simpler to calculate, and clearer to interpret (Ma and Redmond, 1995). Other authors as well have found the metric intuitive and relatively precise.

$$T_p = \frac{P_o - P_r}{1 - P_r} \tag{7-27}$$

$$P_r = \frac{1}{N^2}\sum_{i=1}^{M} x_{i+} n_i \tag{7-28}$$

Where n_i is the *a priori* marginal distribution of the reference data. In other words n_i/N is the *a priori* probability of class membership. While the true value of P_r is found by using Equation 7-28, in practice it is often calculated using Equation 7-29. The latter equation assumes that the column marginals (x_{+i}) are representative of the entire image class proportions. This means a random sampling method is required to generate a confusion matrix to be used with the Tau coefficient.

$$P_r \cong \frac{1}{N^2}\sum_{i=1}^{M} x_{+i}^2 \qquad (7\text{-}29)$$

In the special case where the *a priori* marginal distribution (n_i) is the same for all classes (N/M). This means all classes are equally prevalent.

$$\begin{aligned} P_r &= \frac{1}{N^2}\sum_{i=1}^{M} x_{i+} n_i \\ &= \frac{1}{N^2}\sum_{i=1}^{M} x_{+i}\frac{N}{M} \\ &= \frac{1}{M} \end{aligned} \qquad (7\text{-}30)$$

Substituting the special case of P_r (Equation 7-30) into the standard Tau coefficient formula (Equation 7-27) yields the equal probability form of the Tau coefficient (Equation 7-31).

$$T_e = \frac{P_o - \frac{1}{M}}{1 - \frac{1}{M}} \qquad (7\text{-}31)$$

This equation is the same as the Brennan and Prediger Kappa coefficient (Equation 7-25). In this form, as the number of classes (M) increases the metric will approach the Simple accuracy metric. Both formulations of the Tau coefficient are measurements of the improvement of a classifier over the random assignment of pixels to any class. This quantitative measure has only recently been applied to questions of classification accuracy and is therefore yet to be widely utilized or analyzed by the remote sensing community.

From Equation 7-8 the variance of Simple accuracy (P_o) is already known and the random agreement (P_r) is a constant so its variance is zero. The random agreement parameter is treated as a constant because it is independent of the confusion matrix and can be calculated before classification (Ma and Redmond, 1995).

$$
\begin{aligned}
\sigma^2[T] &= \sigma^2[\frac{P_o - P_r}{1 - P_r}] \\
&= \frac{\sigma^2[P_o - P_r]}{(1 - P_r)^2} \\
&= \frac{\sigma^2[P_o] - \sigma^2[P_r]}{(1 - P_r)^2} \\
&= \frac{P_o(1 - P_o)}{N(1 - P_r)^2}
\end{aligned}
\qquad (7\text{-}32)
$$

Below is a sample Tau coefficient calculation with the appropriate 95% confidence interval for the sample confusion matrix data in Figure 7-2.

$$P_r = \frac{1}{256^2}\left(57^2 + 111^2 + 20^2 + 20^2 + 11^2 + 37^2\right) = 0.272522$$

$$T_p = \frac{0.78125 - 0.272522}{1 - 0.272522} = 0.699 = 69.9\%$$

$$\sigma^2[T] = \frac{0.78125(1 - 0.78125)}{256(1 - 0.272522)^2} = 1.26\text{x}10^{-3}$$

The 95% confidence interval of $T_p = 0.699 \pm (1.96)\sqrt{1.26\text{x}10^{-3}} = (0.630, 0.769)$.

7.4.3 Single Class Metrics

Often the accuracy of an individual class is desired. A metric of the total map accuracy, derived from all the elements in the confusion matrix, gives no indication of the accuracy of any one specific class. Nor does this single metric give any information about the class distribution of the accuracy. Classes frequently exhibit drastically differing accuracies (Story, 1986) and combine to produce misleading overall accuracy metrics. Single class metrics can point out specific problems in a class map, and lead to improved accuracy by indicating which classes should be retrained. Clearly, no single metric could ever fully characterize the quality of a product map due to the lack of information about individual category accuracies.

7.4.3.1 Producer's Accuracy Metric

The simplest, and most commonly used, metric for representing the accuracy of a single class is known as the 'Producer's accuracy'. The Producer's accuracy ($A_p[i]$) is defined in Equation 7-33. It is calculated by dividing the number of correctly classified pixels in a class (diagonal element) by the total number of reference pixels in that class (column marginal).

$$A_p[i] = \frac{x_{ii}}{x_{+i}} \qquad (7\text{-}33)$$

Where: x_{ii} is number of pixels in the i^{th} row and i^{th} column, and x_{+i} the column marginal.

The resulting probability indicates the chance that a randomly selected reference pixel will be correctly classified for the given class. This measures errors of omission because pixels which have not been correctly classified have been omitted from the correct category. This metric is known as the Producer's accuracy because the producer of the classified image is interested in how well a specific area in a scene can be mapped (Story, 1986).

The variance for the Producer's accuracy can be found using Equation 7-34.

$$\sigma^2[A_p(i)] = \frac{1}{N} A_p[i]\left(1 - A_p[i]\right)$$

Where: $N = x_{+i}$ and $A_p[i] = \frac{x_{ii}}{x_{+i}}$

$$= \frac{x_{ii}}{x_{+i}^2}\left(1 - \frac{x_{ii}}{x_{+i}}\right) \qquad (7\text{-}34)$$

7.4.3.2 User's Accuracy Metric

A single misclassification results in a pixel being omitted from the proper class and inadvertently committed into another class. The 'User's accuracy' ($A_u[i]$) metric, also known as 'reliability'(Congalton,1985), is a measure of the probability of an error of commission in a single class. It is calculated by dividing the number of properly classified pixels (diagonal elements) by the total number of pixels classified in that class (row marginal). The resulting quantity is the probability that a pixel from a classified image actually represents that class in the scene. This metric is so named because map user's are interested in how well the map represents what is really in the scene (Story,1986).

$$A_u[i] = \frac{x_{ii}}{x_{i+}} \tag{7-35}$$

The variance for the User's accuracy can be found using Equation 7-36.

$$\sigma^2[A_u(i)] = \frac{1}{N} A_u[i](1 - A_u[i])$$

Where: $N = x_{i+}$ and $A_u[i] = \frac{x_{ii}}{x_{i+}}$

$$= \frac{x_{ii}}{x_{i+}^2}\left(1 - \frac{x_{ii}}{x_{i+}}\right) \tag{7-36}$$

Below are sample calculations of the Producer's and User's accuracies for the sample confusion matrix data in Figure 7-2.

$$A_p[1] = \frac{41}{57} = 0.719 = 71.9\% \qquad A_u[1] = \frac{41}{57} = 0.719 = 71.9\%$$

$$A_p[2] = \frac{92}{111} = 0.829 = 82.9\% \qquad A_u[2] = \frac{92}{105} = 0.876 = 87.6\%$$

$$A_p[3] = \frac{18}{20} = 0.900 = 90.0\% \qquad A_u[3] = \frac{18}{23} = 0.783 = 78.3\%$$

$$A_p[4] = \frac{16}{20} = 0.800 = 80.0\% \qquad A_u[4] = \frac{16}{25} = 0.640 = 64.0\%$$

$$A_p[5] = \frac{5}{11} = 0.454 = 45.5\% \qquad A_u[5] = \frac{5}{14} = 0.357 = 35.7\%$$

$$A_p[6] = \frac{28}{37} = 0.757 = 75.7\% \qquad A_u[6] = \frac{28}{32} = 0.719 = 71.9\%$$

Examples of the User's and Producer's accuracies confidence intervals for the second class are given below.

The Producer's accuracy variance, using Equation 7-34:

$$\sigma^2[A_p(2)] = \frac{92}{111^2}\left(1 - \frac{92}{111}\right) = 1.28\text{x}10^{-3}$$

The 95% confidence interval of $A_p[2] = 0.829 \pm (1.96)\sqrt{1.28\text{x}10^{-3}} = (0.759, 0.899)$.

The User's accuracy variance, using Equation 7-36:

$$\sigma^2[A_u(2)] = \frac{92}{105^2}\left(1 - \frac{92}{105}\right) = 1.03\text{x}10^{-3}$$

The 95% confidence interval of $A_u[2] = 0.876 \pm (1.96)\sqrt{1.03\text{x}10^{-3}} = (0.813, 0.939)$.

8. Discussion

Most of the sampling methods and accuracy metrics which have been discussed suffer from systematic error. This bias has a direction and magnitude. While the direction is often fixed for a given method, there are ways of decreasing its magnitude. A user cannot avoid bias in accuracy assessment due to a variety of real world reasons. What is needed is a strategy for minimizing bias, or at least for balancing optimistic and conservative bias.

8.1 Optimistic Bias

Optimistic bias refers to accuracy assessment methods which result in metrics which overestimate the true classification accuracy. Many methods for assessing classification introduce an optimistic bias but careful treatment can minimize their magnitude. An important step in minimizing optimistic bias is not using training fields or areas near training fields for accuracy assessment. Using dependent data sets, or the selection of reference data that is somehow related to the training data almost always over inflates the measured accuracy (Hammond and Verbyla, 1996). Optimistic bias also commonly results from the restriction of sampling to pixels that are relatively easy to classify. For example, this occurs by selecting relatively homogenous pixels which lie in the center of iso-class regions (Irons *et al*, 1985).

8.2 Conservative Bias

Conservative bias refers to accuracy assessment methods which result in metrics which underestimate the true classification accuracy. There are several common sources of conservative bias, many of which can be avoided. If any class assignment error exists in the reference data then the classification accuracy will appear more in error than it really is. Errors in reference data can occur from any class changes due to temporal offsets between verification and classification. Reference data collection is also open to positional errors due to misregistration which is inherent between the reference and classified data. Another difficulty in obtaining accurate reference data is the difference between the pixel size of the classified image and the minimum mapping unit of reference polygons derived from aerial photography/video data.

Verbyla and Hammond (1995) have analyzed parameters which will lead to metrics that underestimate the true accuracy of class maps derived from various space based sensors. The two parameters of interest were misregistration and the difference in size between MMU and the pixel size in the classified image. These factors are relevant when the confusion matrix is generated using reference data collected from a second image, usually an aerial photograph from an underflight. First the effects of misregistration were analyzed using a SPOT HRV image. The scene was classified using an ISODATA procedure. The classified image was copied and the copy was shifted by one pixel to simulate misregistration. A confusion matrix was then constructed using the classified image as the classified data and the shifted copy as the reference data. This procedure simulates a perfectly classified image with no interpretation error but a small amount of introduced positional error. Shifting in each of the four directions resulted in Tau accuracy values between 64 and 85 percent. Clearly, the accuracy should have been 100 percent but the accuracy was significantly underestimated due exclusively to a single pixel shift. The same procedure was repeated on a Landsat-TM image with similar results. The effects of MMU were also measured. Two copies of the classified image were used to simulate a perfectly classified image with a perfectly coregistered reference image. It was found that as the MMU was increased from one to five times the pixel size, the accuracy ranged from 90 to 48 percent respectively. It was also determined that as the number of class categories is increased, the estimate of classification accuracy will become more conservative. This result is important because it further reinforces the need for careful interpretation of classification accuracy metrics. The measured accuracy will often depend on the number of class categories, as well as other factors, in addition to the true classification accuracy. For example, when comparing two class maps of equal accuracy the measured accuracy of the map with more classes will be lower if any misregistration is present.

8.3 Confusion Matrix Marginal Distribution Scaling

In general, when user selected polygons are utilized as the reference source for accuracy assessment, the size of each polygon is not representative of their true abundance. For example, small reference polygons are often selected for relatively large members. The same is true for reference data collected by stratified random sampling where relatively small class categories are sampled in disproportionately high numbers. However, the accuracy of the Simple, Kappa, Brennan and Prediger's Kappa, and Tau coefficients are contingent upon the proportion of reference samples in each class being the same as their true relative distribution. While the true relative class abundance cannot be known without verification by rigorous, truly random sampling, it can be closely approximated by using the *post priori* probability distribution after classification. In the context of this thesis, the term *post priori* probability will be referring to the proportion of pixels in each class of the final class map. The *a priori* probability will refer to the true class probability of the original image. Confusion matrix marginal distribution scaling by class map *post priori* probabilities can result in more accurate accuracy assessments based on reference data from dependent, independent, and stratified random sampling.

The column marginals of a confusion matrix correspond to abundance of each class category in the reference data. Ideally, these values should be exactly proportional to the true abundance of each class category in the original image. If the final class map abundance is used in place of the truth, the desired confusion matrix marginals can be approximated. This scaling can be accomplished through matrix multiplication of the original confusion matrix by a diagonal matrix with the appropriate scaling coefficients (α_i) along the diagonal as shown in Figure 8-1.

$$\begin{pmatrix} X_{11} & X_{12} & X_{13} & X_{14} & \rightarrow & X_{1m} \\ X_{21} & X_{22} & X_{23} & X_{24} & \rightarrow & X_{2m} \\ X_{31} & X_{32} & X_{33} & X_{34} & \rightarrow & X_{3m} \\ X_{41} & X_{42} & X_{43} & X_{44} & \rightarrow & X_{4m} \\ \downarrow & \downarrow & \downarrow & \downarrow & \searrow & \\ X_{m1} & X_{m2} & X_{m3} & X_{m4} & & X_{mm} \end{pmatrix} \cdot \begin{pmatrix} \alpha_1 & 0 & 0 & 0 & \rightarrow & 0 \\ 0 & \alpha_2 & 0 & 0 & \rightarrow & 0 \\ 0 & 0 & \alpha_3 & 0 & \rightarrow & 0 \\ 0 & 0 & 0 & \alpha_4 & \rightarrow & 0 \\ \downarrow & \downarrow & \downarrow & \downarrow & \searrow & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha_m \end{pmatrix} = \begin{pmatrix} \alpha_1 X_{11} & \alpha_2 X_{12} & \alpha_3 X_{13} & \alpha_4 X_{14} & \rightarrow & \alpha_m X_{1m} \\ \alpha_1 X_{21} & \alpha_2 X_{22} & \alpha_3 X_{23} & \alpha_4 X_{24} & \rightarrow & \alpha_m X_{2m} \\ \alpha_1 X_{31} & \alpha_2 X_{32} & \alpha_3 X_{33} & \alpha_4 X_{34} & \rightarrow & \alpha_m X_{3m} \\ \alpha_1 X_{41} & \alpha_2 X_{42} & \alpha_3 X_{43} & \alpha_4 X_{44} & \rightarrow & \alpha_m X_{4m} \\ \downarrow & \downarrow & \downarrow & \downarrow & \searrow & \\ \alpha_1 X_{m1} & \alpha_2 X_{m2} & \alpha_3 X_{m3} & \alpha_4 X_{m4} & & \alpha_m X_{mm} \end{pmatrix}$$

Figure 8-1 Confusion Matrix Marginal Scaling

The values of the scaling coefficients can be found using Equation 8-1. These coefficients simply scale all the elements along the respective columns of the confusion matrix. The values obtained for α_i can range from zero to infinity. The final scaled matrix will have the same overall sum (N) as the original, unscaled matrix.

$$\alpha_i = \frac{\mathrm{PDF_c}(i)}{\mathrm{PDF_r}(i)} \qquad (8\text{-}1)$$

Where: α_i is the confusion matrix scaling coefficient for column i,
$\mathrm{PDF_c}(i)$ is the probability distribution function of the class map, and
$\mathrm{PDF_r}(i)$ is the probability distribution function of the reference data.

To determine the values of the scaling coefficients, the discrete probability distribution functions (PDF) of the class map (Equation 8-2) and the reference data (Equation 8-3) must first be found. It should be noted that what is referred to as discrete probability distribution functions in this thesis are often known as probability mass functions in more formal contexts. The class map PDF, also known as the *post priori* probability distribution, is found by normalizing the image wide histogram. Pixels which are left undefined by the classifier are not included.

$$\mathrm{PDF_c}(i) = \frac{c_i}{H - U} \qquad (8\text{-}2)$$

Where: c_i is the total number of pixels in the class map belonging to class i,
H is the total number of class map pixels, and
U is the total number of undefined pixels in the class map.

The reference data PDF is found by simply dividing the column marginals by the total number of elements in the confusion matrix.

$$\mathrm{PDF_r}(i) = \frac{x_{+i}}{N} \tag{8-3}$$

Where: x_{+i} the column marginal, and
N is the number of pixels in the matrix.

Like all PDF's, the class map and reference PDF's should sum over all classes (Equation 8-4) to exactly one.

$$\sum_{i=1}^{M} \mathrm{PDF}(i) = 1.0 \tag{8-4}$$

Confusion matrix marginal distribution scaling is useful in cases where the class distribution of the reference data is different from that of the true distribution. It assumes that the class map probability distribution is at least, a more accurate estimator of the true class distribution than the reference marginals. Scaling a matrix which utilizes randomly sampled reference data will not result in a gain, but can introduce unnecessary error. However, with non-proportionally sampled reference data, matrix scaling will obtain a more accurate assessment. Unlike normal accuracy assessment, this technique requires more information than just a confusion matrix. The final class map is also needed to determine its histogram. However, the class map must already be available to generate the confusion matrix. Marginal distribution scaling is convenient and lends itself well to automation because no additional input is required from the user.

Multiplication of an entire confusion matrix by a constant does not change the value of any accuracy coefficient calculated based on that matrix. It does, however, change the value of the confidence interval of any metric. Scaling by value greater than one reduces the range of a confidence interval at a fixed level of significance. Marginal distribution scaling multiplies each column in a matrix by a different value. These values are chosen in such a way (Equation 8-1) that the overall matrix sum is preserved. In this way, still assuming an adequate number of samples, the confidence intervals of the all accuracy metrics are still valid. In general, column marginal distribution scaling will change the value of the Simple, Kappa, Brennan and Prediger's Kappa, Tau, and User's accuracy coefficient metrics. Scaling the columns has no effect on the Weighted or Producer's accuracy coefficients. However, due to round-off error, the Weighted and Producer's accuracy coefficient metrics often varied slightly between the scaled and unscaled matrices. The more pixels contained in the confusion matrix, the less noticeable this difference becomes.

8.3.1 Example of Confusion Matrix Scaling

An example has been provided to demonstrate confusion matrix marginal distribution scaling. Figure 8-2 is the class map for the example. It has dimensions 15 x 15 pixels so there are a total (*H*) of 225 pixels and 18 of which were left unclassified (*U*).

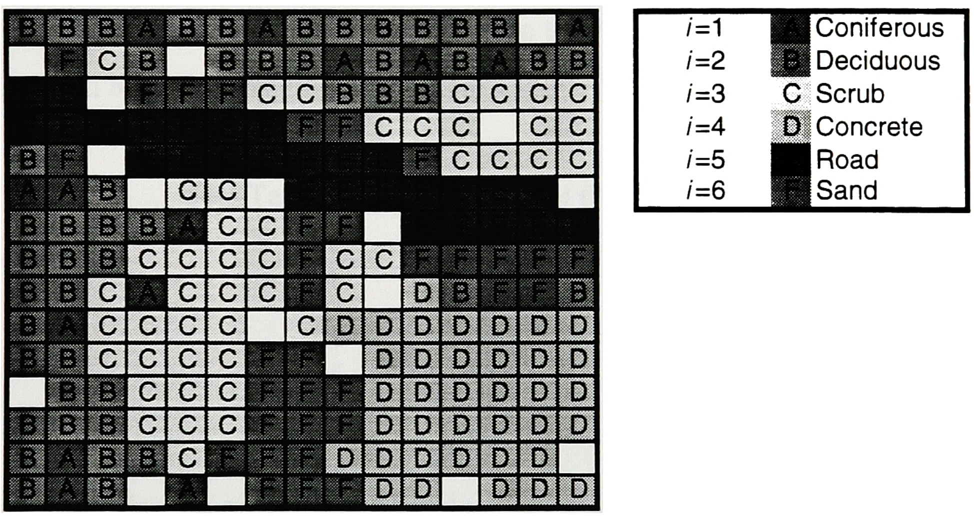

Figure 8-2 Sample Class Map

The class map was then verified using stratified random sampling with an equal number of samples in each strata (*M*) for a total of 24 samples (*N*). The resulting confusion matrix is shown in Figure 8-3. The original class map and the number of accuracy assessment samples is unrealistically low only for the purpose of providing a convenient example.

Reference Data

Classified Data

17	Coniferous	Deciduous	Scrub	Concrete	Road	Sand	$i+$
Coniferous	3	1	0	2	0	0	6
Deciduous	0	2	0	0	1	0	3
Scrub	0	1	4	1	0	0	6
Concrete	1	0	0	1	0	0	2
Road	0	0	0	0	3	0	3
Sand	0	0	0	0	0	4	4
$+i$	4	4	4	4	4	4	24

Figure 8-3 Confusion Matrix for Sample Class Map

The PDF's for the class map and reference data are then calculated. The PDF for the class map is calculated based only on the pixels which were classified using Equation 8-2. The PDF for the reference data is derived from the confusion matrix column marginals using Equation 8-3. The results are shown in Table 8-1 and Table 8-2.

Table 8-1 Distribution of Class Map

i	c_i	$PDF_c(i)$
1	14	0.068
2	48	0.232
3	47	0.277
4	37	0.179
5	28	0.135
6	33	0.159

The scaling coefficients are then calculated using Equation 8-1. The results of this calculation are also posted in the last column of Table 8-2.

Table 8-2 Distribution of Reference

i	x_{+i}	$PDF_r(i)$	α_i
1	4	0.167	0.406
2	4	0.167	1.391
3	4	0.167	1.362
4	4	0.167	1.072
5	4	0.167	0.812
6	4	0.167	0.957

Finally, in Figure 8-4, the original confusion matrix is multiplied by the scaled diagonal matrix. The resulting matrix, however, is not comprised exclusively of integer values as is the case with traditional confusion matrices. Integer values are not necessary to evaluate any of the accuracy metrics or their confidence intervals.

$$\begin{pmatrix} 3 & 1 & 0 & 2 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 \\ 0 & 1 & 4 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 \end{pmatrix} \cdot \begin{pmatrix} 0.406 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.391 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.362 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.072 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.812 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.957 \end{pmatrix} = \begin{pmatrix} 1.218 & 1.391 & 0 & 2.144 & 0 & 0 \\ 0 & 2.782 & 0 & 0 & 0.812 & 0 \\ 0 & 1.391 & 5.448 & 1.072 & 0 & 0 \\ 0.4060 & 0 & 0 & 1.072 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.436 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3.828 \end{pmatrix}$$

Figure 8-4 Scaling Sample Confusion Matrix

If desired, the scaled confusion matrix can be rounded as shown in Figure 8-5. However, rounding introduces unnecessary round-off error in the calculation of accuracy metrics. The sum of the scaled matrix in Figure 8-4 is 24, the same as the original confusion matrix. The sum of the rounded scaled matrix is 22. The source of this discrepancy is round-off error as well.

Reference Data

Classified Data

16	Coniferous	Deciduous	Scrub	Concrete	Road	Sand	i+
Coniferous	1	1	0	2	0	0	4
Deciduous	0	3	0	0	1	0	4
Scrub	0	1	5	1	0	0	7
Concrete	0	0	0	1	0	0	1
Road	0	0	0	0	2	0	2
Sand	0	0	0	0	0	4	4
+i	1	5	5	4	3	4	22

Figure 8-5 Scaled Confusion Matrix for Sample Class Map

8.3.2 Kolmogorov-Smirnov Testing of Post Priori Probabilities

The benefits of confusion matrix marginal distribution scaling by *post priori* probabilities are contingent upon an accurate approximation of the true class distribution. The Kolmogorov-Smirnov two-sample test was used to determine the fitness of class map *post priori* probability distributions for this purpose. The Kolmogorov-Smirnov test is a nonparametric statistic which verifies or rejects the hypothesis of equality between two cumulative probability distributions. The synthetic images utilized in this thesis allow a quantitative comparison between the two distributions of interest because the exact true distribution is known.

The Kolmogorov-Smirnov test is a goodness-of-fit procedure which starts with random samples from two unknown distributions and makes no assumptions about their shape. For the purpose of testing the fitness of class map histograms for the estimation of the true class distribution, both matched pair by class quantities are treated as discrete PDF's. The true class distribution is known for the synthetic images. Next, the cumulative probability function (CDF) of the class map and true distribution can be found from their PDF's using Equation 8-5.

$$\mathrm{CDF}(i) = \sum_{x=1}^{i} \mathrm{PDF}(x) \qquad (8\text{-}5)$$

The null hypothesis (Equation 8-6) for the Kolmogorov-Smirnov two sample test states that two CDF's come from the same or identical populations with respect to location and dispersion. In this case, this would indicate that the class map distribution is a good indicator of the true class distribution. The alternative hypothesis (Equation 8-7), that two CDF's differ, would indicate that the class map is a poor predictor of the true class distribution.

$$H_0\text{: } \mathrm{CDF_c}(i) = \mathrm{CDF_T}(i) \text{ for all values of } i \qquad (8\text{-}6)$$

$$H_1\text{: } \mathrm{CDF_c}(i) \neq \mathrm{CDF_T}(i) \text{ for at least one value of } i \qquad (8\text{-}7)$$

The two sided test statistic, used to accept or reject the null hypothesis, is given by Equation 8-8. It is equal to the supremum, over all x, of the absolute value of the difference between two cumulative distribution functions (Wayne, 1990).

$$D = \mathrm{Max}[\ |\mathrm{CDF_c}(x) - \mathrm{CDF_T}(x)|\ ] \text{ for all values of } x \qquad (8\text{-}8)$$

Table 8-3 lists the test statistic (D) threshold at which the null hypothesis is rejected at the ninety percent confidence level for a given number of samples (n). In this case, the number of samples corresponds to the number of classes.

Table 8-3 Quantiles of the Smirnov Two Sample Test Statistic of size n

n	P=0.90
3	$2/3$
4	$3/4$
5	$3/5$
6	$4/6$
7	$4/7$
8	$4/8$
9	$5/9$
10	$5/10$
11	$5/11$
12	$5/12$

Source: Birnbaum and Hall, 1960

The Kolmogorov-Smirnov two sample, two sided test threshold can also be approximated with $\frac{1.73}{\sqrt{n}}$ for sample sizes (n) of greater than forty at a ninety percent level of significance.

When the test statistic is less than or equal to the value in the table, there is no evidence to reject the null hypothesis, so the two samples are presumed to be from the same distribution. Test statistic values greater than the table suggest rejecting the null hypothesis in favor of the alternative.

The *forest* scene, due to its synthetic nature, is ideally suited, along with the Kolmogorov-Smirnov test, for testing the viability of *post priori* probabilities as an estimator of the true class probability distribution. The true class map probability distribution is known for this special case, along with the *post priori* probabilities. Using the Kolmogorov-Smirnov test, these quantities can be compared.

The *post priori* probabilities were calculated for nine *forest* scene class maps, from the Gaussian Maximum Likelihood, Rule Based Genetic Algorithm (Mystic), and Fuzzy ARTMAP classifiers with image visibility's of 5km, 7km, and 23 kilometers. All nine images originated from one synthetic scene, so they all had the same true class probability distribution. In Figure 8-6, the *post priori* distribution of the GML classified, 23km visibility image [$PDF_c(i)$] is compared to the true distribution.

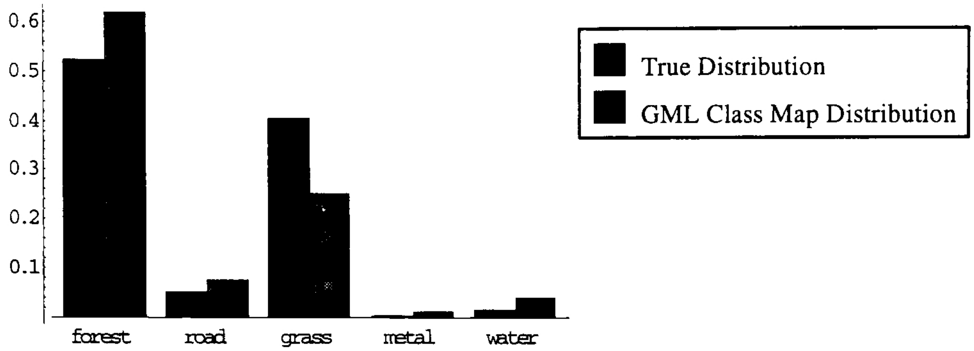

Figure 8-6 Forest Class Probability Distributions

In Figure 8-6, the cumulative *post priori* distribution [$CDF_c(i)$] is compared to the true cumulative class distribution [$CDF_T(i)$].

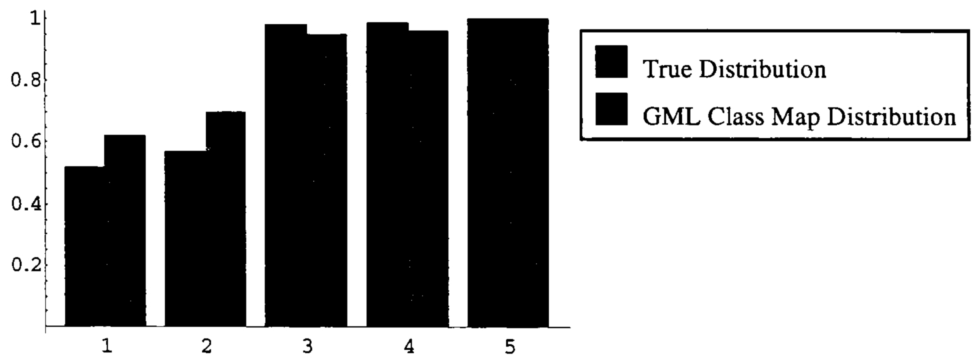

Figure 8-7 Forest Class Cumulative Probability Distributions

From these cumulative distributions, the Kolmogorov-Smirnov test statistic (Equation 8-8) was calculated. The test statistic was found by selecting the largest absolute difference (Table 8-4) between the two cumulative distributions.

Table 8-4 Calculation of Smirnov Test Statistic for Forest 23k GML Class Map

i	$CDF_c(i)$	$CDF_T(i)$	$\lvert CDF_c(i)-CDF_T(i)\rvert$
1	0.621	0.522	0.099
2	0.696	0.574	0.122
3	0.951	0.981	0.030
4	0.967	0.984	0.017
5	1.000	1.000	0.000

For the GML classified, 23km visibility image the Kolmogorov-Smirnov test statistic (D) is 0.122.

$D = \text{Max}[\ \lvert CDF_c(i) - CDF_T(i)\rvert\ ] = 0.122$

Since this value is less than the 0.600 threshold (Table 8-3) for 5 samples, there was no evidence to reject the null hypothesis.

The same test was repeated for the other eight *forest* scene images. The results of these tests are summarized in Table 8-5.

Table 8-5 Results of Kolmogorov-Smirnov Two-Sample Test

	Mystic	GML	Fuzzy ARTMAP
5km Visibility	$CDF_c(x)=CDF_T(x)$	$CDF_c(x)=CDF_T(x)$	$CDF_c(x)=CDF_T(x)$
7km Visibility	$CDF_c(x)=CDF_T(x)$	$CDF_c(x)=CDF_T(x)$	$CDF_c(x)=CDF_T(x)$
23km Visibility	$CDF_c(x)=CDF_T(x)$	$CDF_c(x)=CDF_T(x)$	$CDF_c(x)=CDF_T(x)$

In each case, the null hypothesis was not rejected. These results support that, in fact, *post priori* probabilities are good estimators of true class distributions.

The Kolmogorov-Smirnov test is known to be a relatively permissive test. In practice, the null hypothesis is rarely rejected. This fact is attributable, primarily to two reasons, not an underlying problem with the test. First, some variance normally occurs and is expected in any distribution. Comparing two samples which come from the same distribution can result in observations which appear quite different. Second, the decision regarding the null hypothesis is made based on only two incomplete distributions. This is not a significant amount of observations, which is normally used in statistical hypothesis testing. The ability of the Kolmogorov-Smirnov test to accurately discriminate between the same or two different distributions increased as more samples (n) are taken from each. A stricter test could be made by employing a different approach such as clustering analysis.

9. Results

The purpose of this study was to examine several factors involved in classification accuracy assessment. Two particular questions were researched. First, to what extent do real world imaging parameters, specifically spatial resolution and atmospheric visibility, degrade classifier performance. The second, how is classification accuracy assessment best accomplished. The answer to these questions depends on how the classification accuracy is measured. Classification accuracy assessment involves making an educated estimate of entire class map accuracy based on incomplete and often biased data. For this reason, the problem of measuring the effect of stressing factors on classifier performance cannot be completely separated from the method of accuracy measurement. Both problems were studied with equal detail as part of this task. The findings of this work are presented in this section.

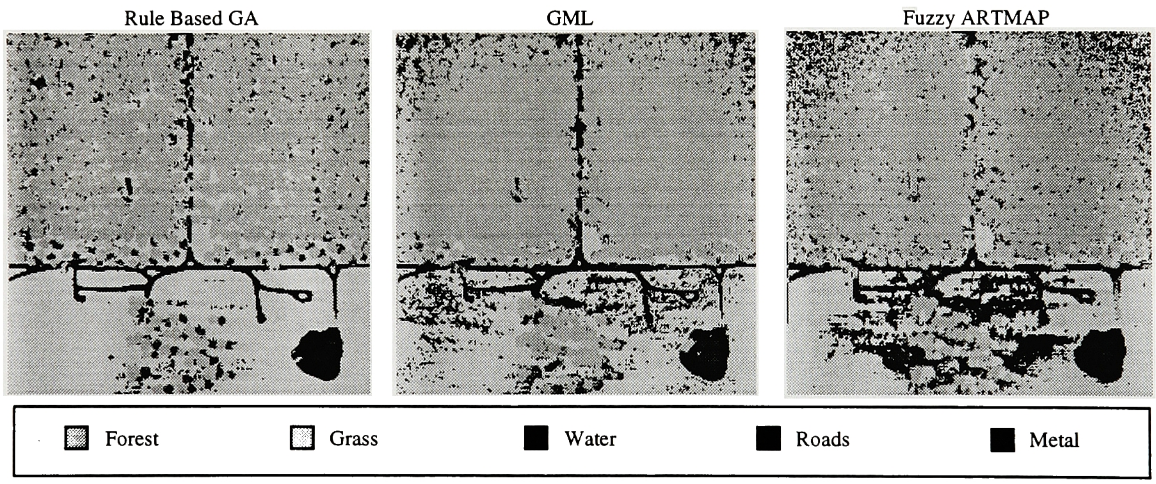

Figure 9-1 Class Maps from Forest 23km Visibility Scene Classification

The original three baseline scenes were degraded in terms of spatial resolution and atmospheric visibility to produce nine images. These images were then classified using the GML, Mystic™, and Fuzzy ARTMAP classifiers.

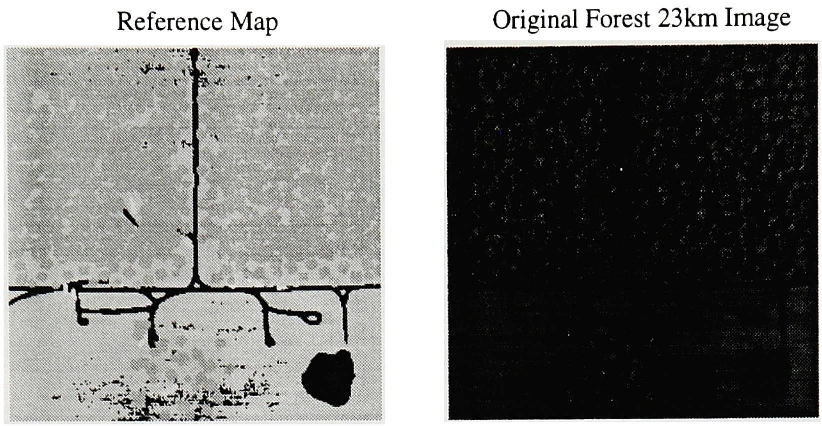

Figure 9-2 Synthetic Reference Map and Original Forest Image

The resulting class maps were evaluated using dependent, independent, randomly sampled, and sample and population synthetic material maps as reference sources. A total of one hundred and nineteen (119) confusion matrices were generated. The simple percent correct, Weighted accuracy, Kappa, Brennan and Prediger's Kappa, and Tau coefficients were calculated and appropriate confidence intervals were provided for each confusion matrix. These results are all contained in Appendix A.

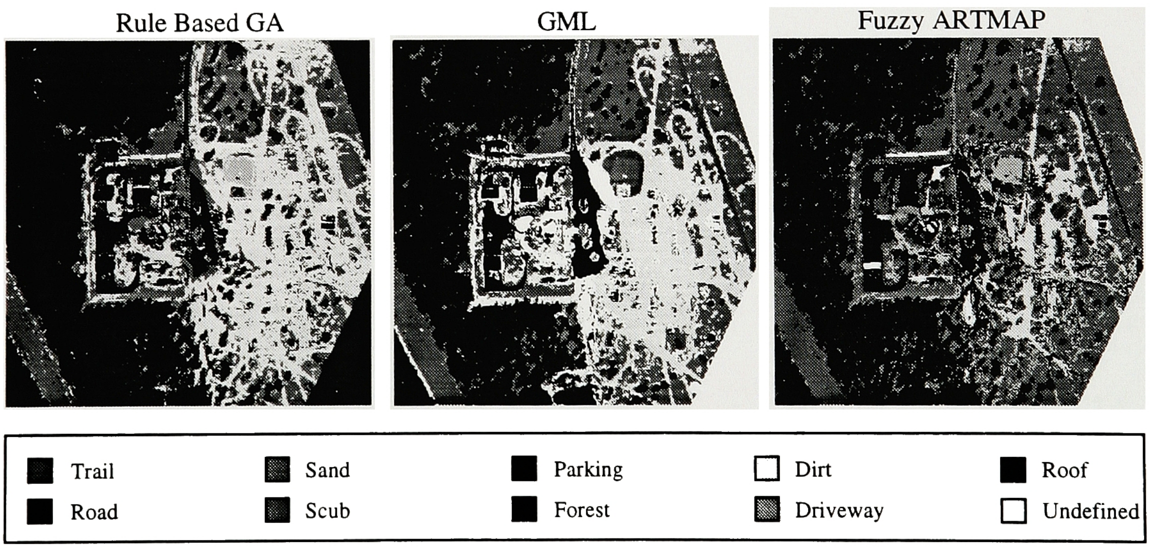

Figure 9-3 Class Maps from Tank 23km Visibility Scene Classification

The Appendix A contains the classification accuracy assessment reports for all the class maps evaluated as part of this work. Each assessment report indicates the scene name, stressing parameters, classification procedure, confusion matrix, and corresponding accuracy metrics with confidence interval. The confusion matrices have reference data along the columns and classified data along the rows. Scenes with the same resolution but different atmospheres share common training and reference data. Images with difference resolution required different training and reference data at each of the three resolutions.

Three sample class maps for the *forest* scene are shown in Figure 9-1. Each color represents one of the five class categories which are listed in the key below. Each class map was classified with the classifier listed above. The original synthetic scene, for which the class maps are shown, had atmospheric visibility of 23 kilometers and a GIFOV of 1 meter. For purposes of visual comparison, the original image and true class map are provided in Figure 9-2.

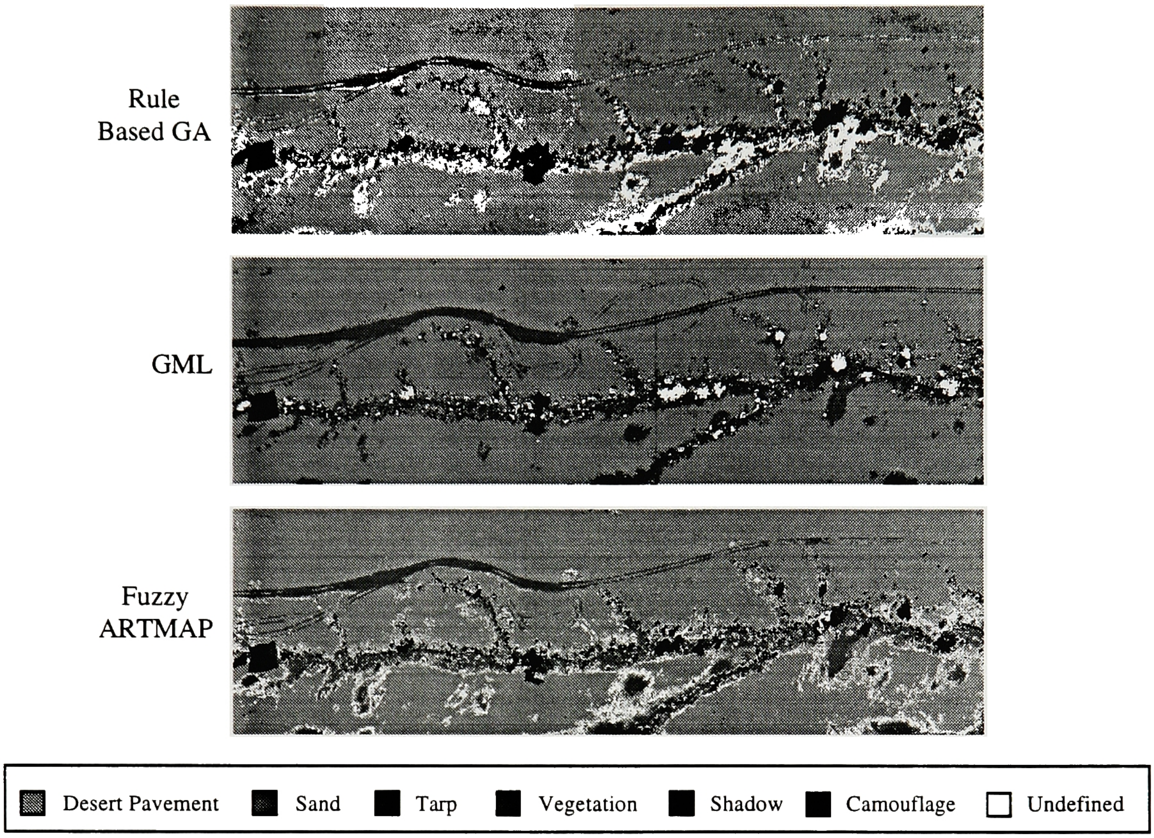

Figure 9-4 Class Maps from Desert 1m GIFOV Scene Classification

Three of the *tank* scene class maps are shown for the 23 kilometer visibility with 1 meter GIFOV image in Figure 9-3. This scene was classified with nine categories. The corresponding original image is shown in Figure 5-5. Three of the *desert* scene class maps are shown in Figure 9-5. The original image had a 1 meter GIFOV and a clear atmosphere. This scene was trained with six categories. The corresponding original image is shown in Figure 5-6.

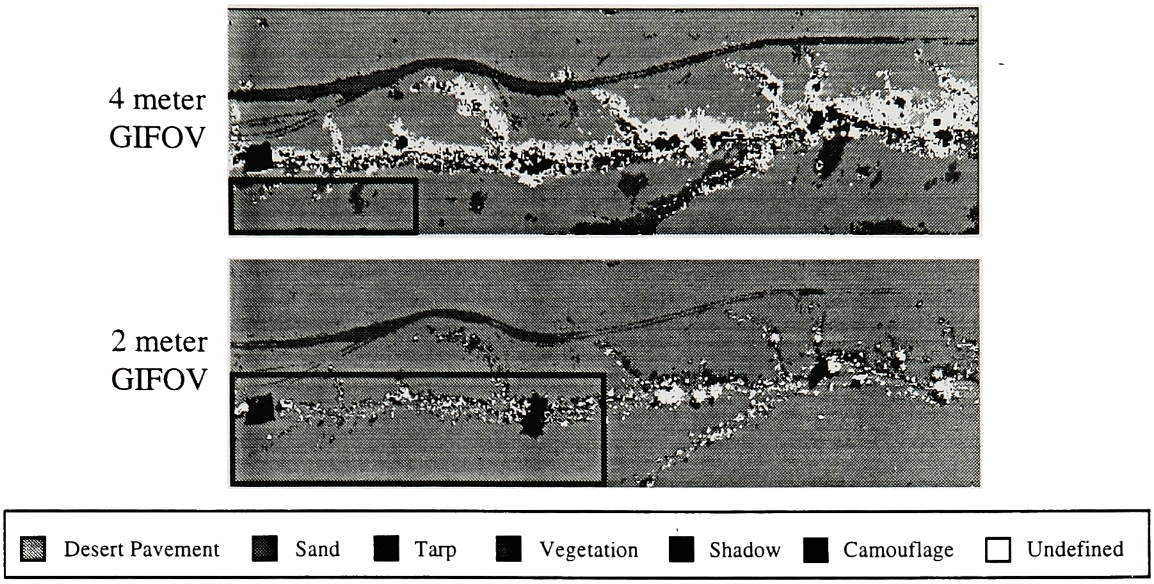

Figure 9-5 Class Maps from Desert 2m & 4m GIFOV Scene GML Classification

The same desert scene was also degraded in terms of spatial resolution. Two class maps from degraded imagery with 4 meter and 2 meter GIFOV are shown in Figure 9-5. A rectangle is superimposed over each image to indicate the relative size of each.

9.1 Effect of Reference Data Source

First, the results from the analysis of the effect of reference data source on assessed classification accuracy will be discussed. Obtaining high quality reference data is the most difficult aspect of assessment. Many classification product maps claiming high accuracy have been evaluated with reference data derived from biased sources. Accuracy metrics do not account for inaccuracy or biased reference data gathered from improper sampling. The results of this work have verified that more accurate assessment procedures require more time and expense. There are no quick fixes. For example, randomly gathered reference data has constantly been proven to yield more representative samples, and therefore it is more accurate than larger, user selected regions. This added fidelity, however, comes at a cost. The dependent set, such as the one on the left hand side of Figure 7-4, required no additional time to gather other than what was required to train the classifier. The independent reference set for the corresponding image required only fifteen minutes to gather. However, the most robust method, pure random sampling, required almost six hours for the *tank* scene. The 278 points, shown in Figure 9-6, required approximately one minute and fifteen seconds per point.

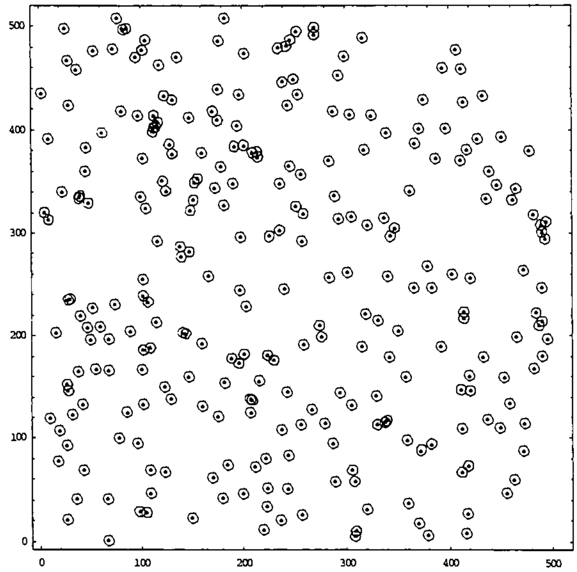

Figure 9-6 Random Sampling of 278 Points

Table 9-1 was compiled to demonstrate typical accuracy assessment results based on different reference data sources. In this table, the assessed accuracy of three class maps is indicated, each using three different reference sources. The table is based on GML classification of the *forest* scene and uses the simple percent correct accuracy metric. The accuracy coefficients in the synthetic reference column are the exact accuracies for the entire scene and therefore serve as an image wide truth for examining the other two reference sources. This same summary is also shown for the *tank* and *desert* scenes in Table 9-2 and Table 9-3 respectively.

Table 9-1 Effect of Forest Reference Source on Measured Percent Correct

	Dependent	Independent	Synthetic
23km Visibility	0.943 (3)	0.913 (3)	0.807 (2)
7km Visibility	0.963 (2)	0.978 (2)	0.865 (1)
5km Visibility	0.983 (1)	0.984 (1)	0.772 (3)

Clearly, based on the data provided in Table 9-1, the accuracy assessments based on dependent and independent reference sets overestimate the true class map accuracies. This small sampling showed the results obtained using dependent and independent data to be very similar to each other, but not a very good indicator of the true class map accuracy. In fact, the rank ordering (noted in parenthesis) of the image accuracy's is not even maintained using the first two reference sources. Accuracy metrics evaluated using these two reference sources will be of uncertain precision and therefore are inappropriate for serious accuracy assessment projects requiring the absolute accuracy.

Table 9-2 Effect of Tank Reference Source on Kappa Coefficient

	Dependent	Independent	Random
23km Visibility	0.996 (3)	0.987 (1)	0.850 (1)
7km Visibility	1.000 (1)	0.982 (3)	0.827 (2)
5km Visibility	0.997 (2)	0.983 (2)	0.827 (2)

Table 9-2 summarizes the Kappa values obtained by assessing the *tank* scene class maps using various sources of reference data. The class maps for each level of atmospheric visibility were generated using the GML classifier. Table 9-3 shows the effect of reference data sources on the measured Tau coefficient value for the *desert* scene. Each Tau value is shown with its rank ordering and corresponding scene resolution and reference source. All images were classified using the Mystic™ classifier. The trend of these two tables is the same in the *forest* scene.

As expected, accuracy metrics which were calculated based on dependent reference data sets were higher than those calculated with independent and random reference data for the same class map. Independent reference data also resulted in higher accuracy metric values than those based on random reference sources. In addition, the difference between the dependent and independent values was small compared to the difference between the independent and random values. Due to the low corresponding metric values and high statistical integrity of random sampling it is clear that dependent and independent reference sources both yield overly optimistic results. This overly inflated nature was also apparent in comparisons against synthetic reference. A large portion of bias encountered with user selected data is attributable to its poor ability to predict *a priori* probabilities.

Table 9-3 Effect of Desert Reference Source on Tau Coefficient

	Dependent	Independent	Random
1m GIFOV	0.856 (1)	0.796 (1)	0.614 (1)
2m GIFOV	0.771 (2)	0.717 (2)	0.573 (2)
4m GIFOV	0.675 (3)	0.632 (3)	0.557 (3)

In Figure 9-7, a comparison between accuracy assessment reference data sources is made. In this figure the Simple accuracy metric is plotted for assessments of the *desert* scene based on dependent, independent, and random reference data. The scenes were all classified with the GML classifier and the results are shown for resolutions of 1, 2, and 4 meters. This plot shows that the trend is consistent. The dependent reference overestimates the independent which overestimates the random reference accuracy assessment. The trend of the dependent reference assessment is not even correct. If an analysis of GIFOV was based only on this dependent data, the investigator would be mislead into believing the 2 meter image had higher classification accuracy than the 1 meter GIFOV image.

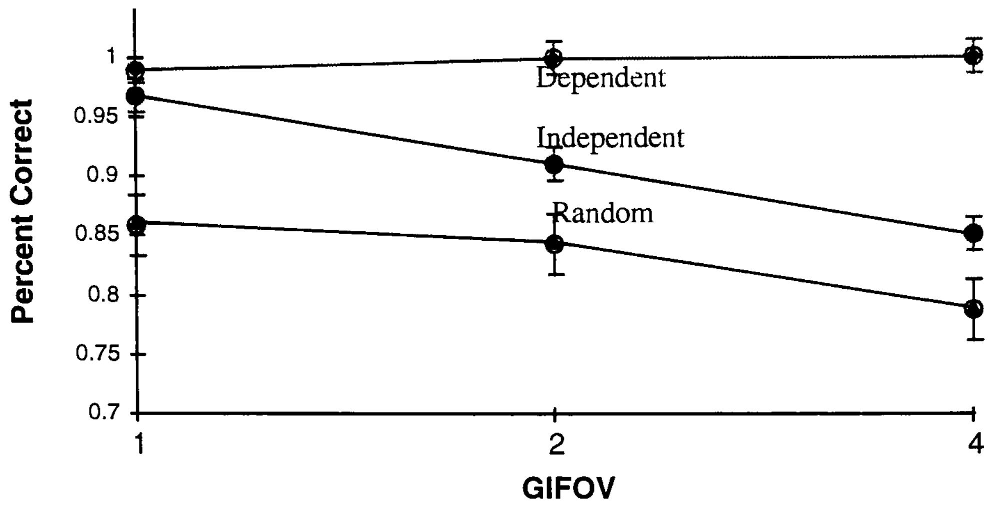

Figure 9-7 Multisource Assessment of Desert Scene GML Class Map

The Simple accuracy coefficient obtained using random reference was compared to their true values found using synthetic reference. The comparison was made on the forest scene using all three classifiers and levels of atmospheric visibility. The results are shown in Figure 9-8. The nine bars in this chart show the difference between the random reference percent correct minus the synthetic reference percent correct for each image. In seven out of nine cases, the random accuracy was larger than the true accuracy. However, this was true by a small amount, and is attributable to pure chance. In fact, the random accuracy was within 6 percent of the true value for each case. In addition, the true answer was within the random reference data, Simple accuracy metric 95% confidence interval for eight out of the nine images.

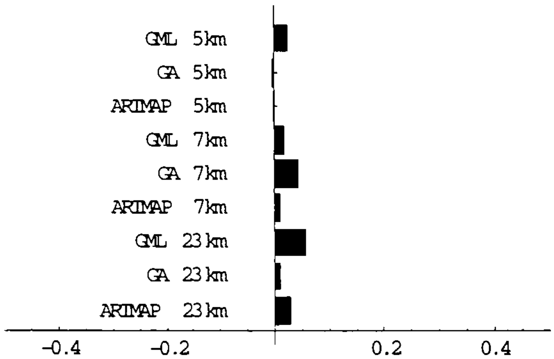

Figure 9-8 Error of Random Forest Assessment

Table 9-4 and Table 9-5 are confusion matrices, but not normal matrices generated based on a class maps. These tables are provided to demonstrate one of the problems with user selected reference. Table 9-4 is a confusion matrix with synthetic reference from the *forest* scene on the columns and the dependent reference set from the *forest* scene along the rows. The confusion matrix in Table 9-5 has synthetic reference from the *forest* scene on the columns and the independent reference data from the *forest* scene along the rows.

Table 9-4 Verification of Dependent Reference Data

	forest	road	grass	metal	water	
forest	523	0	53	0	0	576
road	0	170	51	0	0	221
grass	0	4	273	0	0	277
metal	0	2	0	26	0	28
water	0	0	4	0	98	102
	523	176	381	26	98	

During accuracy assessment, reference data, by its very nature, is expected to be completely correct. No accuracy metrics, or conventional confidence intervals account for mistaken reference data. However, the user selected reference data for the *forest* scene could be verified because the complete synthetic truth was available. In fact, the dependent reference was wrong 9.5 percent (114/1204) of the time and the independent reference data was incorrect for 3.3 percent (90/2712) of the pixels. The grass and road classes accounted for the majority of confusion. The large polygons used to identify user selected reference frequently promote the selection of a few incorrect pixels.

Table 9-5 Verification of Independent Reference Data

	forest	road	grass	metal	water	
forest	268	0	2	0	0	270
road	0	153	30	0	0	183
grass	0	47	1830	0	0	1877
metal	0	1	9	70	0	80
water	0	0	1	0	301	302
	268	201	1872	70	301	

When an image analyst interactively selects reference regions for the purpose of accuracy assessment, such as the case with dependent and independent data sets, there will be many sources of inaccuracy. Reference regions which are completely typical of the entire image are crucial to unbiased accuracy assessment because this incomplete data must serve as an estimator of the entire image. In general, regions selected by a user will not be representative of the entire image. User selected regions tend to avoid particularly heterogeneous image areas, however, these are the same areas where classifiers often fail and assessment is most desirable. In other words, the high spatial correlation of user selected regions are inversely correlated with the highly spatially correlated regions of classifier error. Dependent data, used to train the classifier, has an additional disadvantage. By design, classifier performance is optimized in these regions to produce the least error. These are not typical areas and will inevitably yield optimistically biased accuracy assessment.

Another inaccuracy introduced by user selected regions is their poor estimation of *a priori* probabilities. Most accuracy metrics required confusion matrix reference marginals which are equivalent to the original image class proportions. In general, the size of user selected regions bear no resemblance of the true abundance of each class in the original image. The unpredictable and inaccurate results obtained with the dependent and independent data sets in this task are attributed to this fact. The results of confusion matrix scaling by *post priori* probabilities to correct for this problem will be discussed later is this section.

9.2 Accuracy Metric Results

The final product of classification accuracy assessment is usually a single accuracy coefficient including its confidence interval. However, single coefficient accuracy metrics cannot fully characterize classification accuracy. One image wide metric cannot completely convey the accuracy of small, and possibly important, class categories while maintaining proper emphasis on larger classes. Accuracy metrics such as the Simple accuracy, Weighted, Kappa, Brennan and Prediger's Kappa and Tau coefficients, however, can indicate an overall quality of a given class map or classifier. A useful metric can be defined as a single, intuitive index of accuracy ranging from 0 to 1.0 which measures a meaningful quantity. Most of the metrics evaluated as part of this task fall under this definition but it is necessary to interpret each one appropriately.

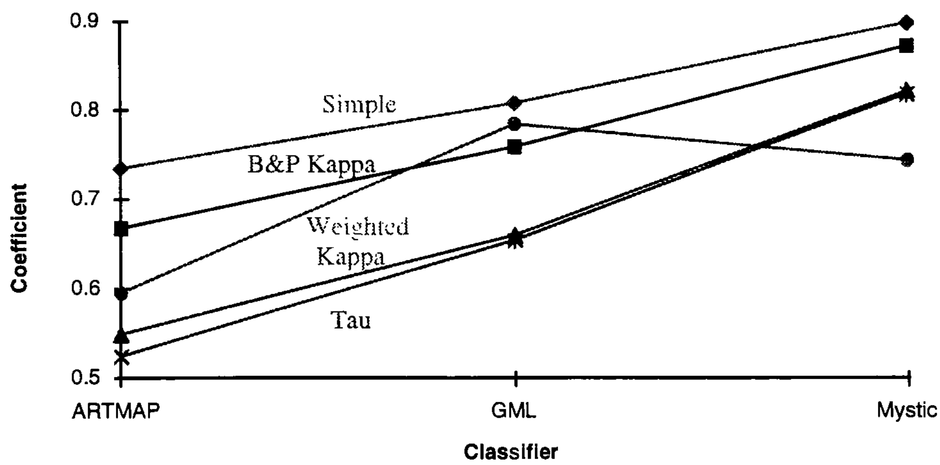

Figure 9-9 Classifier Performance on Forest 23km 1m Image

Figure 9-9, based on synthetic reference data, compares the five accuracy metrics for one image classified by all three classifiers. This scene is unique in that it is the only case where Mystic™ outperformed the GML classifier. As expected, a clear trend was observable in the rank ordering of the classification accuracy metrics. The rank ordering of the accuracy metrics in this figure were typical, with the exception of the Weighted accuracy metric. With most confusion matrices, the Simple accuracy coefficient yielded the highest value. In the far majority of cases it was followed by Brennan and Prediger's Kappa, Kappa, and lastly the Tau coefficient. Occasionally, the Brennan and Prediger's Kappa and standard Kappa resulted in identical values. The rank ordering of the Weighted accuracy metric was not consistent. It ranged from the lowest to the highest.

The Kappa, Brennan and Prediger's Kappa and the Tau coefficient all measure essentially the same quantity but utilize three different approaches for calculating it. They attempt to measure the percent of properly classified pixels due only to classifier performance. This is done by subtracting a coefficient of chance agreement from the percent of properly classified pixels. The Kappa, Brennan and Prediger's Kappa, and Tau coefficients all have the same general form given by Equation 9-1.

$$P = \frac{P_o - A}{1 - A} \tag{9-1}$$

Where: P is the respective coefficient value,
P_o is the percent correct coefficient, and
A is the respective chance agreement term.

The three metrics do, however, differ in the way they measure the coefficient of chance agreement (A). The Kappa uses the confusion matrix joint marginal distributions (Equation 7-20) to find the chance agreement term while the Tau coefficient uses only the column marginal distribution (Equation 7-29). The Brennan and Prediger's Kappa coefficient assumes the marginal distributions are fixed and uses a chance term equal to the reciprocal number of classes. The respective chance agreement coefficients for all three metrics are summarized in Table 9-6.

Table 9-6 Chance Agreement Coefficients

Kappa	Brennan and Prediger's Kappa	Tau
$A = \frac{1}{N^2}\sum_{i=1}^{M} x_{i+}x_{+i}$	$A = \frac{1}{M}$	$A = \frac{1}{N^2}\sum_{i=1}^{M} x_{+i}^2$

The metrics are normalized by (1-*A*), so the coefficient *P* range from 0 to 1.0 under realistic conditions. For a fixed percent accuracy, the Kappa, Brennan and Prediger's Kappa, and Tau coefficients will decrease when the chance agreement terms are increased.

Often, authors report only one metric when detailing class map accuracies. This causes difficulties if other researchers wish to compare the accuracy of class maps which have each been assessed with different metrics. The entire confusion matrix is needed to calculate the Simple, Weighted, Kappa, Brennan and Prediger's Kappa, Tau, Producer's, and User's accuracy coefficients. Given only one metric it is impossible to convert the value to any other metric with the exception of the Simple and Tau coefficients. A conversion between these two metrics can be made without the entire confusion matrix as long as the number of class categories is known. Converting to the Tau from the Simple coefficient is accomplished by using the normal Tau formula (Equation 7-31). The conversion to the Simple from the Tau coefficient can be found by the inverse of the Tau formula (Equation 9-2).

$$P_o(T_e, M) = \frac{1 - T_e}{M} + T_e \qquad (9\text{-}2)$$

Where: T_e is the Tau Coefficient, and
M is the number of classes.

In addition, the Weighted accuracy can be found if the Producer's accuracy is known for each of the classes.

9.2.1 Simple Accuracy

The simple percent correct accuracy coefficient is easy to calculate and interpret. It serves as the basis for calculating several other statistical metrics which utilize marginal distributions, but it is also useful in its own right. The Simple accuracy metric yields higher values than the Brennan and Prediger's Kappa, the Kappa, and the Tau for a particular reason. The Simple accuracy metric measures the proportion of pixels properly classified. The other three metrics, using similar but different approaches, measure the proportion of pixels properly classified minus the number which were properly classified by chance. These two different types of metrics are really answering two different questions. The first method answers the question about the absolute accuracy of a given class map. The second method answers the question about how well a given classifier has performed. The Simple accuracy is useful for end users of class maps who want to know how accurate they are. The Simple accuracy metric assumes the confusion matrix column marginal distribution is indicative of the true image-wide class distribution. This is not the case, in general, with user selected reference regions. Use of the Simple accuracy metric should be restricted to cases where truly random or *a priori* stratified random sampling is used to gather reference data.

9.2.2 Weighted Accuracy

The Weighted accuracy metric is not a highly intuitive metric and does not yield a clear interpretation. These factors greatly limit its practicality as a useful indicator of classification accuracy. It is rarely used in practice and barely meets the criteria set forth by the definition of a useful metric. The Weighted accuracy is the only metric which does not assume the confusion matrix marginals are equal to the *a priori* probabilities. Therefore, it is the only metric who's assumptions are not violated by using dependent and independent reference data or any other sampling method. The Weighted accuracy metric assumes each class category is of equal importance. It possibly could be applicable in special cases were the relative size of class categories is not known or of interest.

9.2.3 *Kappa Coefficient*

The Kappa coefficient is a categorical data analysis index of agreement between two data sets, which is used when chance agreement is a concern. Like the Brennan and Prediger's Kappa and the Tau coefficient, it attempts to measure the classifier accuracy while subtracting the portion that were correct due only to chance agreement. The Tau coefficient assumes that the product of the confusion matrix row and column marginals are equivalent to the *a priori* probabilities. This metric is therefore not valid with user selected training regions. In the majority of cases encountered in this study, the Kappa value was greater than the Tau and slightly smaller than the Brennan and Prediger's Kappa.

It is the opinion of the author that the Kappa coefficient can be a misleading metric for representing the accuracy of a class map. For example, imagine classifying an image that resulted in the following 2 x 2 confusion matrix.

$$\begin{pmatrix} 10 & 11 \\ 0 & 10 \end{pmatrix}$$

In this case 20 out of a total 31 pixels were properly classified. The corresponding Simple and Kappa coefficient values are 0.65 and 0.37 respectively. Now suppose the same image was classified again and resulted in this confusion matrix.

$$\begin{pmatrix} 11 & 5 \\ 5 & 10 \end{pmatrix}$$

In this case, one of the 11 pixels which had been misclassified in the first classification is properly classified. In addition, the marginal distributions have been rearranged. Now, 21 out of a total of 31 pixels have been properly classified. The corresponding Simple and Kappa coefficients are 0.68 and 0.35. Clearly, the second classification was more accurate than the first. The Simple accuracy metric reflects this increase with a 0.03 increase in the metric value. However, if the Kappa coefficient was used as an index it would appear as though there was actually a decrease in class map accuracy. The Kappa coefficient decreased 0.02 while the class map accuracy truly increased. In this circumstance, the Kappa coefficient does not even maintain ordinal scaling in relation to the true class map accuracy. This statement is based on the probable assumption that a matrix with more diagonal elements is considered more accurate than a matrix with less diagonal elements and the same total number of pixels. These factors indicate that the Kappa coefficient is a poor choise for analyzing the effect of stressing parameters such as MTF or atmospheric visibility. For instance, a Kappa comparison could report a decrease when a process in fact increased the classification accuracy.

While a small matrix was utilized for the ease of illustration, the same phenomena can occur for larger confusion matrices. The reason for the misleading nature of the Kappa coefficient is the way it uses the joint marginals to arrive at the chance agreement term. Confusion matrices which have clustered off-diagonal, misclassified pixels will result in higher Kappa values than matrices with evenly distributed off-diagonal elements. For example, take the following confusion matrix:

$$\begin{pmatrix} 10 & 10 \\ 0 & 10 \end{pmatrix}$$

In this matrix, all the off-diagonal elements are grouped in one location. The Kappa value is 0.40. In the following matrix the same number of pixels are properly classified, but the marginals are evenly distributed.

$$\begin{pmatrix} 10 & 5 \\ 5 & 10 \end{pmatrix}$$

This gives a Kappa coefficient of 0.33. The Kappa indicates a measured decrease in accuracy of 0.07 units while in actuality the class map quality was unchanged. This fact indicates that the Kappa coefficient cannot serve as a interval scale for comparing class map accuracy's.

In addition to its misleading nature, it is also the opinion of the author that the Kappa coefficient values are counterintuitive. The Kappa coefficient spans a range from -1.0 to 1.0. Most confusion matrix accuracy coefficients range from 0.0 to 1.0 and are often interpreted as probabilities or percentages. The Kappa is zero when the elements are all the same and negative when the number of off-diagonal terms is greater than the number of diagonal elements. Therefore, when the Simple accuracy for a given confusion matrix is less than 0.50, the Kappa can be, but is not necessarily, negative.

$$\begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}$$

For instance the above matrix has a Kappa value of -0.20.

For many applications in categorical data analysis, the Kappa coefficient is a useful statistic. However, it has been shown to be poorly suited for representing class map accuracy. While the Kappa coefficient has been accepted by many authors in the remote sensing community, other metrics, notable the proportion of properly classified pixels, offer significant advantages. Many authors have championed the Kappa coefficient because it utilizes all the information in a confusion matrix. This is true because it uses both marginal distributions to calculate the chance agreement term. This fact does not seem useful to class map users who are not interested in the marginal distribution of errors.

9.2.4 Brennan and Prediger's Kappa

The Brennan and Prediger's Kappa is similar to the standard Kappa but it uses a smaller chance agreement term. Using a smaller chance agreement term explains its highest rank ordering when compared against the Kappa and Tau coefficients. Brennan and Prediger's Kappa coefficient does not use the confusion matrix marginals to calculate its chance agreement term. Therefore, no assumptions are violated by employing user selected reference data. This metric, however, makes the erroneous assumption that all classes are equally prevalent. The Brennan and Prediger's Kappa chance agreement term depends only on the number of class categories and not marginal distributions. Therefore, it does maintain the same rank ordering as the Simple accuracy metric.

9.2.5 Tau Coefficient

The Tau coefficient measures the improvement in chance agreement between two data sets over random assignment. In the context of classification accuracy assessment, the improvement in chance agreement is due to the digital image classifier. When applied to the analysis of a confusion matrix, the Tau coefficient, is equal to the proportion of pixels in the matrix which are properly classified (along the diagonal) due only to classifier performance. The metric does not count the proportion of pixels which are properly classified due only to chance agreement. The Tau coefficient assumes that the confusion matrix column marginals are equivalent to the *a priori* probabilities, so only randomly sampled reference is applicable. The Tau coefficient consistently resulted in the lowest value.

Like the Kappa, the Tau coefficient is not an intuitive metric. It does not result in a ordinal scale compared to the proportion of properly classified pixels. The Tau behaves in the opposite manner of the Kappa coefficient in that its value decreases when off-diagonal elements of the confusion matrix are clustered rather than equally distributed. Like Kappa, an equal number of pixels in each category of the confusion matrix results in a zero Tau value. It spans a range from negative infinity to unity.

9.3 Effect of Stressing Parameters

Many factors such as the classifier used, number of classes, time of day, and signal to noise ratio effect the ultimate classification accuracy. In these sections, the results of two such parameters, image resolution and atmospheric visibility, are presented.

9.3.1 Resolution

Low resolution images, with large ground instantaneous fields of views (GIFOV), usually result in higher classification accuracy. High resolution imagery is more difficult to classify accurately. This result has been verified by several researchers (Gong and Howarth, 1990). Low resolution imagery, due to the central limit theorem, tends to have a more Gaussian spectral probability distribution. The non-Gaussian nature of high resolution imagery should therefore favor non-parametric classifiers such as Mystic™ and the Fuzzy ARTMAP. However, this result was not verified in this task. High resolution imagery was used in this study and the parametric classifier still outperformed the non-parametric classifiers.

This study utilized aerial imagery to examine the effect of resolution on classification accuracy. The *desert* scene was used at three different GIFOV's. The limiting factor in allowing the analysis of the effect of spatial resolution was the original spatial dimensions of the images. Only the *desert* scene was large enough to allow 4x4 convolution with down sampling and still be large enough to be classified. The GML classifier was unable to accurately classify any images smaller that 250x250 pixels. This factor required that the original image be no smaller that 1000x1000 pixels. In the future, utilizing satellite imagery in addition to aerial imagery would avoid this problem.

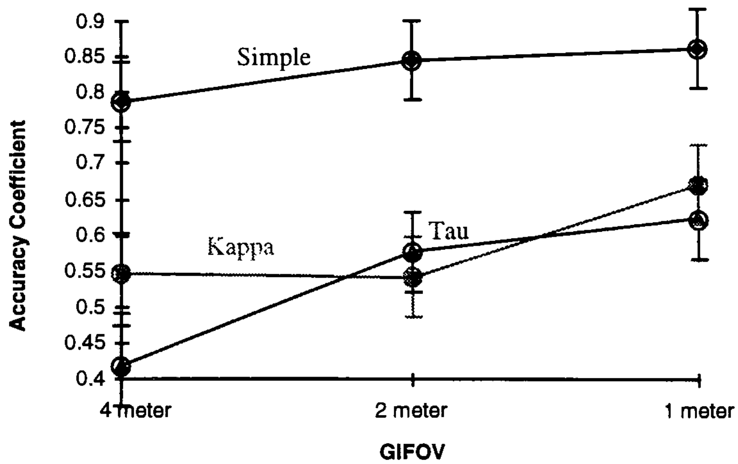

Figure 9-10 GML Classification Accuracy for Desert Scene

Plotted in Figure 9-10 is a summary of effect of spatial resolution on GML classification accuracy of the *desert* scene. The results are shown for each of the three spatial resolutions. The assessment was based upon randomly sampled reference data. The three metrics are provided to help visualize the trend. As can easily be seen, classification accuracy actually increased with higher spatial resolution. This is the opposite of what was expected. However, it is interesting to note, that when GIFOV was changed from 4 to 2 meters the Simple and Tau coefficient increased while the Kappa coefficient decreased. This supports the premise put forth in the last section that the Kappa coefficient is a misleading metric. If the analysis of the effect of GIFOV on classification accuracy was made solely using the Kappa statistic, it would yield differing results.

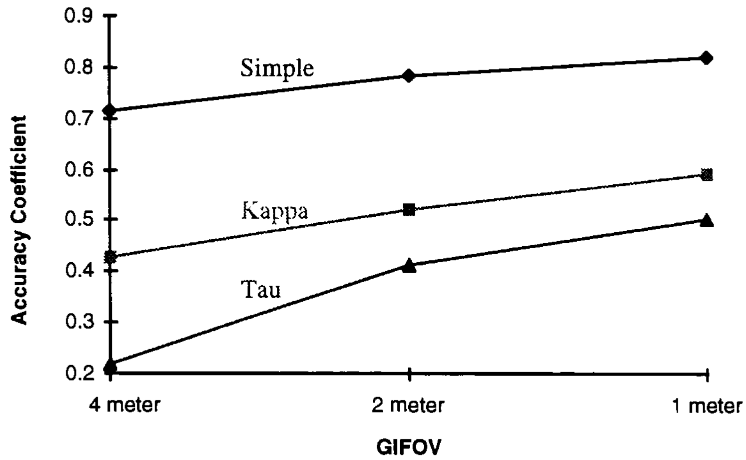

Figure 9-11 ARTMAP Classification Accuracy for Desert Scene

Plotted in Figure 9-11 is a summary of effect of spatial resolution on Fuzzy ARTMAP classification accuracy of the *desert* scene. The results are shown for each of the three spatial resolutions. The assessment was based upon randomly sampled reference data. Again, classification accuracy actually increased with higher spatial resolution. For this classifier all metrics increased accordingly with higher image spatial resolution.

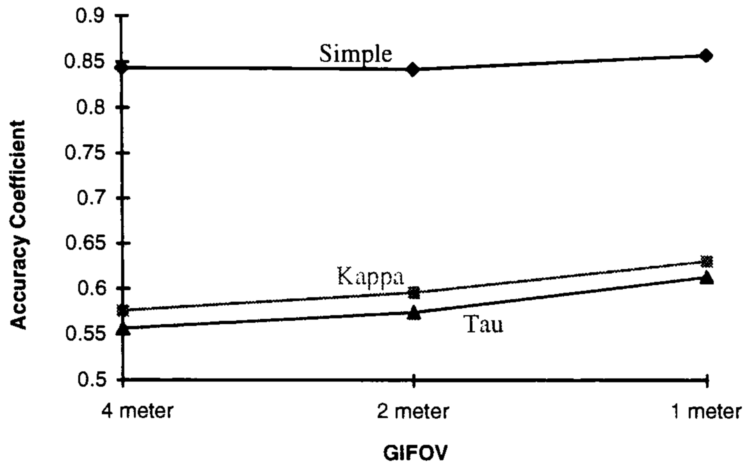

Figure 9-12 Mystic™ Classification Accuracy for Desert Scene

Finally, in Figure 9-12 the effect of spatial resolution on Mystic™ classification accuracy of the *desert* scene is plotted. The results are again shown for each of the three spatial resolutions. The assessment was based upon randomly sampled reference data. Again, with this classifier higher spatial resolution appeared to result in higher classification accuracy assessment. This trend is broken in only one case. The Simple accuracy metric resulted in a slight decrease in classification accuracy between the 4 and 2 meter GIFOV.

The results obtained regarding the effect of spatial resolution on classification accuracy in the section are the opposite of what was expected. However, the results of this study are consistent and independent of accuracy metric or classifier. The dependent reference data did occasionally provide inconsistent results which will be considered in err. The trends observed can be explained by the very high resolution of the aerial imagery employed in this study. Usually, satellite imagery is used to verify that classification accuracy is increased with lower resolution. In these cases, the move toward spectrally Gaussian shaped classes aids classifier performance. In the *desert* scene, high resolution image, the classes were spectrally distinct. In addition, due to the nature of this scene, feature classes such a bushes, shadows, and camouflage were very small, often only a few pixels across. The far majority of pixels were spectrally pure pixels. At lower resolution, many pure pixels were converted to mixed pixels. This blurred the spectral distinction between classes and made classification more difficult. This phenomena was observed directly during classifier training. It became more difficult to select homogeneous training regions on low resolution images.

9.3.2 Atmosphere

Atmospheres for this task were simulated using two different approaches. The first approach was taken using the *tank* scene. Three versions of this scene were generated, one with a 23km, 7km, and 5km visibility. The atmospheres were simulated using a linear histogram operation that reduced the scene contrast. The result was atmospheric simulations that best simulated a sensor with a narrow field of view (FOV). This type of atmosphere had little effect on overall classification accuracy. For example, the GML classifier obtained Kappa values of 0.850, 0.827, and 0.827 respectively with randomly sampled reference data.

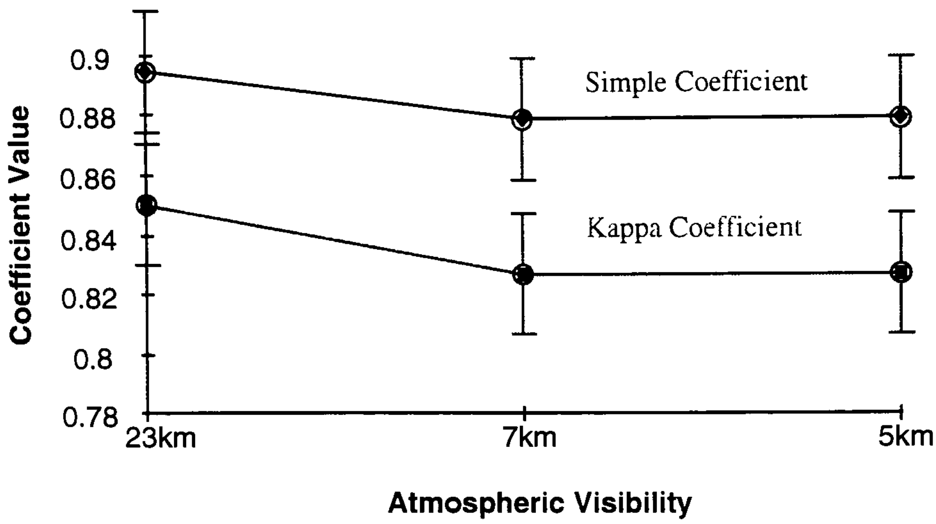

Figure 9-13 Effect of Tank Scene Atmospheric Visibility on Classifier Performance

The plot in Figure 9-13 summarizes the effect of the simulated atmospheric visibility on the GML classification accuracy of the *tank* scene. The results are shown in terms of the Simple and Kappa coefficient accuracy metrics. The effect of visibility was not dramatic. In fact, the accuracies did not change between the 7 and 5 kilometer visibilities. The way in which the *tank* scene's atmosphere was simulated did not account for the increased path length off nadir. This adds a spatial dependence to images levels which great diminishes classifier accuracy as will be seen with the *forest* scene.

The second approach to simulating atmosphere was used on the *forest* scene. This scene was generated with 23km, 7km, and 5km visibility as well. This method for simulating the atmosphere was similar to the first but it also accounted for increased path length off nadir as well. This highly realistic approach best models a sensor with a wide field of view.

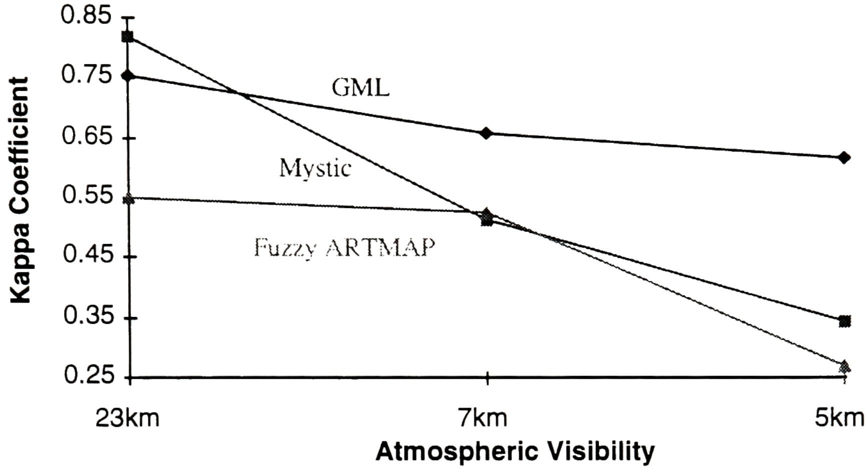

Figure 9-14 Effect of Forest Scene Atmospheric Visibility on Classifier Performance

Classifier performance degraded quickly with decreasing visibility in the *forest* scene as seen in Figure 9-14. The Kappa values in Figure 9-14 were evaluated using synthetic reference data. The GML classifier proved to be the least sensitive to atmospheric visibility and Mystic™ proved to be the most sensitive. In fact, when the atmosphere was very heavy, with a 5km visibility, Mystic™ lost the ability to discriminate the road class. Heavy atmospheres also increased the spatial correlation of classifier error for all classifiers. Error rate increases as angle from nadir increases, as can be seen by comparing Figure 9-15 to Figure 9-16.

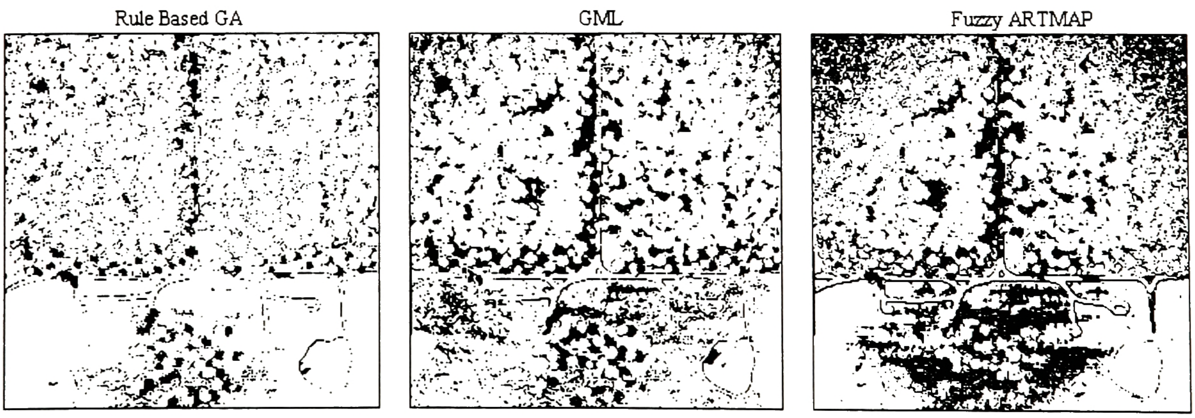

Figure 9-15 Spatial Correlation of Classifier Error at 23km Visibility

Figure 9-15 demonstrates the high spatial correlation of classification error for all three classifiers. Errors are more frequently clustered rather than isolated occurrences. The light colored areas are regions where the classifier properly classified the *forest* 23km visibility scene and the dark areas are regions where pixels were misclassified. These images are possible because the scene was synthetically generated with DIRSIG and reference data is therefore available for the entire image.

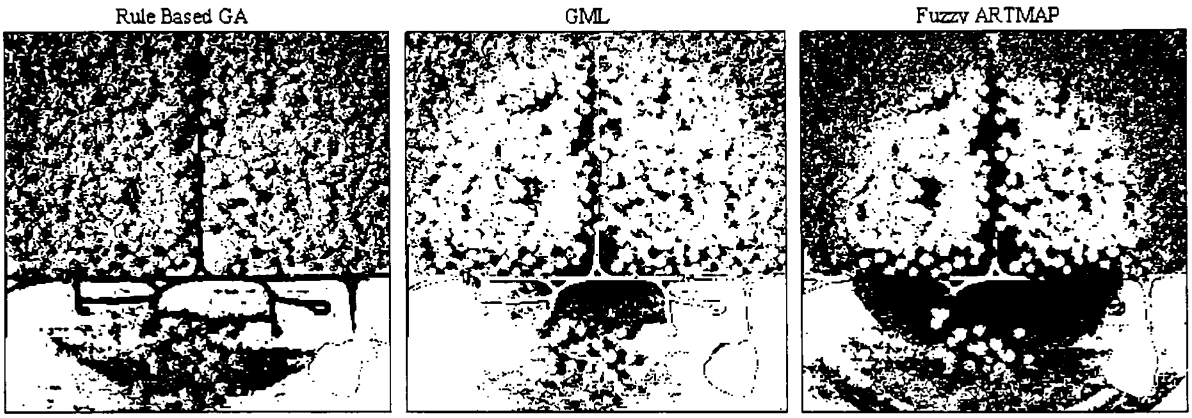

Figure 9-16 Spatial Correlation of Classifier Error at 5km Visibility

9.4 Results of Confusion Matrix Scaling

In this section, the results from correcting for the inaccurate class distribution of user selected reference data are presented. Confusion matrices from all three scene were scaled by their *post priori* probability distributions. All confusion matrices were based on independent reference data. Accuracy metrics calculated from the scaled matrices are then compared to the unscaled and true accuracies metric values.

9.4.1 Scaling of Forest Scene Confusion Matrices

Nine *forest* scene confusion matrices were scaled by their *post priori* probabilities. The original confusion matrices which were scaled were based on an independent reference assessment of the scene classified by the GML, Mystic, and Fuzzy ARTMAP classifiers, with visibility's of 5km, 7km, and 23 kilometers.

The probability distribution of the user selected reference, dependent and independent, along with the true class probability distribution is given in Table 9-7. The exact true distribution is known for this scene because it was synthetically generated.

Table 9-7 Probability Distributions of Reference Data

	Dependent	Independent	Truth
forest	0.478	0.100	0.522
road	0.184	0.068	0.051
grass	0.230	0.692	0.407
metal	0.023	0.030	0.003
water	0.085	0.111	0.016

The *post priori* probabilities (PDF_c), from each of the class maps, is shown in Table 9-8. These values, along with the reference data marginal distribution (PDF_r), were used to calculate the scaling coefficients (α_i).

Table 9-8 Probability Distributions of Forest Scene Class Maps

	GML 23km	GML 7km	GML 5km	GA 23km	GA 7km	GA 5km	ART 23km	ART 7km	ART 5km
forest	0.621	0.627	0.557	0.517	0.749	0.418	0.560	0.645	0.473
road	0.075	0.054	0.144	0.034	0.007	0.000	0.131	0.135	0.317
grass	0.255	0.298	0.260	0.386	0.220	0.292	0.273	0.160	0.157
metal	0.010	0.005	0.023	0.048	0.009	0.013	0.000	0.000	0.000
water	0.038	0.015	0.015	0.015	0.015	0.278	0.036	0.060	0.053

Root mean square (RMS) error can be used to compare the *post priori* probability distribution of the reference data marginal distribution against the true class distribution. These RMS errors were calculated using Equation 9-3.

$$\text{RMS Error} = \frac{1}{M}\sum_{i=1}^{M}\sqrt{\left[\text{PDF}(i) - \text{PDF}_r(i)\right]^2} \qquad (9\text{-}3)$$

Where: M is the number of class categories,
PDF(i) is the probability distribution function, and
$\text{PDF}_r(i)$ is the reference probability distribution function.

The results of these calculations, along with the true, Simple percent accuracy of the corresponding class maps are summarized in Table 9-9. The percent accuracy assessment is based on a complete sample of the synthetic material map.

Table 9-9 RMS Error of Marginal Distribution Approximation for Forest Scene

	RMS Error x100	Percent Accuracy
Dependent	8.84	
Independent	16.9	
GML 23km	6.08	80.7
GML 7km	4.40	86.5
GML 5km	5.89	77.2
GA 23km	1.77	89.7
GA 7km	9.27	75.5
GA 5km	10.9	56.7
ARTMAP 23km	5.48	73.5
ARTMAP 7km	10.0	72.0
ARTMAP 5km	12.1	50.8

If two reference data sets are of equal bias, then the data set with the reference marginal distribution which more closely approximates the true class distribution will result in a more accurate accuracy assessment. Table 9-9 indicates that the probability distribution for all three classifiers, classifying scene at all three levels of atmospheric visibility, resulted in less error then the independent reference set in the case of the *forest* scene. The class map probability proved to be more accurate than the dependent reference source in five of the nine classes. The dependent reference data set used for this image had a RMS error which was lower than would normally be expected. The class maps which yielded worse results than the dependent set had relatively low percent accuracies of below 76 percent correct. In addition, because the Fuzzy ARTMAP classifier failed to resonate with the proper number of classes, the metal category marginal was equal to zero. In general, class map probability distribution will better approximate *post priori* probabilities than reference data marginals.

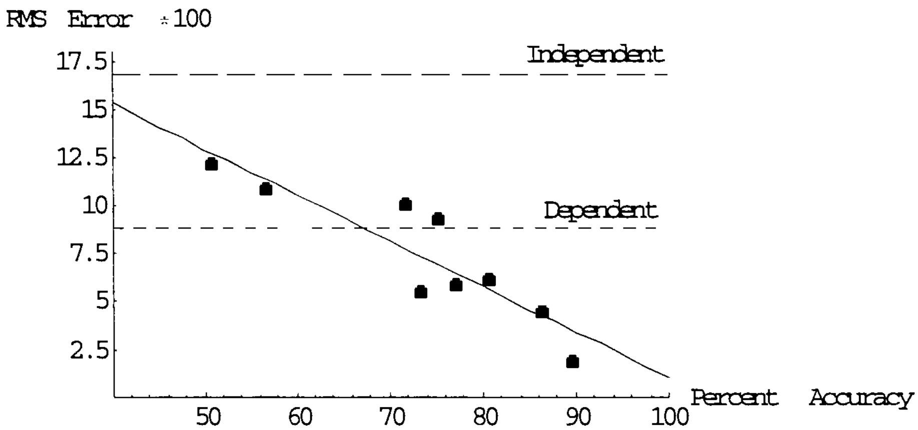

Figure 9-17 Effect of Class Map Accuracy on Distribution Estimation

The percent of correctly classified pixels does not necessary indicate how well the class map probabilities will resemble the true *post priori* probabilities. A class map can have a low percent accuracy and still obtain the correct probability distribution. Classification accuracy assessment measures not only the probability distribution, but also verifies that each class is in the proper spatial location. However, Table 9-9 indicates that there is a strong negative relationship between class maps percent accuracy and RMS error between the class map probability distribution and the true distribution. In Figure 9-17 the user selected reference and *post priori* RMS error is graphically compared to the class map percent accuracy. Each point on the plot represents a particular class map from the *forest* scene classified by one of the three classifiers at one of the three levels of atmospheric visibility. The linear least-squares fit is provided as a solid trend line between these points. The corresponding RMS error of the dependent and independent reference sets is shown as the small and large dashed lines respectively. The linear least-squares fit visually demonstrates the strong negative linear relationship over a range of 50.8 to 89.7 percent accurate between RMS error and percent accuracy, even with three different classifiers used. This fact is reinforced by a large goodness of fit as indicated by a correlation coefficient of -0.89. This result is important in image area estimation, where class map histograms are used to determine percentages of ground cover in large and/or remote geographical regions. These results, however, were obtained using a single image and only two reference sets. This does not prove that the same trends will be observed with all images.

A new set of probability scaled confusion matrices was generated for the *forest* scene based on the independent reference data. Nine matrices corresponding to the GML, Mystic, and Fuzzy ARTMAP classification of the *forest* scene with 5km, 7km, and 23 kilometer visibility were scaled. Confusion matrix marginal distribution scaling was used on the *forest* scene in particular, because the true accuracy of this synthetic scene is already known. The matrices were scaled by the scaling coefficients (α_i) in Table 9-10.

Table 9-10 Scaling Coefficients for Forest Scene

	5km Visibility			7km Visibility			23km Visibility		
	GML	Mystic	ARTMAP	GML	Mystic	ARTMAP	GML	Mystic	ARTMAP
α_1	5.60	4.20	4.75	6.30	7.52	6.48	6.24	5.20	5.62
α_2	2.14	0	4.70	0.81	0.11	2.01	1.11	0.51	1.94
α_3	0.38	0.42	0.23	0.43	0.32	0.23	0.37	0.56	0.39
α_4	0.78	0.43	0	0.18	0.29	0	0.35	1.61	0
α_5	0.14	2.50	0.48	0.14	0.14	0.54	0.35	0.13	0.32

The results of the scaling, compared with those of the unscaled, and true accuracies are show in Table 9-11. The results are reported in the percent of properly classified pixels. The true accuracy is based on complete sampling of the synthetic material map.

Table 9-11 Percent Accuracy Results of Matrix Scaling Forest Scene

		True Accuracy	Independent Reference Accuracy	Scaled Independent Reference Accuracy
5km Visibility	GML	0.772	0.984	0.985
	Mystic	0.567	0.694	0.810
	ARTMAP	0.508	0.617	0.622
7km Visibility	GML	0.865	0.978	0.989
	Mystic	0.755	0.642	0.891
	ARTMAP	0.720	0.579	0.905
23km Visibility	GML	0.807	0.913	0.967
	Mystic	0.897	0.962	0.950
	ARTMAP	0.735	0.819	0.931

The absolute difference between the scaled independent accuracy and the true accuracy, along with the absolute difference between the unscaled independent and true accuracy is summarized in Table 9-12.

Table 9-12 Absolute Error Result of Matrix Scaling Forest Scene

		Independent Reference	Scaled Independent Reference
5km Visibility	GML	0.212	0.213
	Mystic	0.127	0.243
	ARTMAP	0.109	0.114
7km Visibility	GML	0.113	0.124
	Mystic	0.130	0.136
	ARTMAP	0.141	0.185
23km Visibility	GML	0.106	0.160
	Mystic	0.065	0.053
	ARTMAP	0.084	0.196

Unfortunately, the scaled results had more error than the unscaled independent reference in eight out of nine cases. However, these results are easily explainable and do not indicate a problem with matrix scaling in general. User selected reference is an inaccurate predictor of image wide statistics for more than one reason. The probability distribution of user selected reference has already been shown to be inaccurate. In addition, reference sets of this type are usually over optimistically biased and unrepresentative of the image being sampled. Confusion matrix scaling by *post priori* probabilities corrects only for the former problem. The poor results after scaling suggest bias is the overriding source of error in independent reference sets.

In this case specifically, the increase in error introduced by scaling the independent reference confusion matrices is due primarily to bias in the forest class of the scene. On average, across all nine independent confusion matrices, the forest class category had the highest Producer's accuracy. These values were much greater than their true value, observed using synthetic reference data. However, this category only contributed 10 percent of total sum of the independent confusion matrix. After matrix scaling, the forest class contributed on average approximately 50 percent of the total matrix sum, close to its true value of 52.2 percent. This caused an apparent increase in accuracy because the overly positively biased forest class displaced other, lower accuracy classes such as the grass category.

9.4.2 Scaling of Tank Scene Confusion Matrices

Five *tank* scene confusion matrices were scaled by their *post priori* probabilities. The original confusion matrices were based on an independent reference assessment of the scene classified by the GML classifier with 5km, 7km, and 23km visibility's and Mystic™ and Fuzzy ARTMAP classifiers with visibility's of 23 kilometers. The image histograms, used to calculate the *post priori* probability distribution, are listed in Table 9-13.

Table 9-13 Histograms of Tank Scene Class Maps

	GML 23km	GML 7km	GML 5km	GA 23km	ARTMAP 23km
trail	7720	9900	7895	29484	0
sand	44551	43508	47435	42973	17639
parking	6070	6260	6294	9285	12288
dirt	1412	1430	1450	2694	23320
roof	1304	1361	1357	1855	1538
road	6982	8422	8250	8722	22006
scrub	64526	65498	65166	54995	79322
forest	87456	90860	91687	110913	85009
driveway	2596	1639	2707	1223	6274
TOTAL	222617	228878	232241	262144	247396

The user selected reference regions and class map *post priori* distributions were compared against the randomly sampled probability distribution using RMS error. The results are tabulated in Table 9-14. It has already been shown that random sampling results in highly accurate assessments. Random reference data will be assumed to be unbiased for use as a comparison against user selected reference, specifically independent reference sets.

Table 9-14 RMS Error of Marginal Distribution Approximation for Tank Scene

	RMS Error x100
Dependent	9.85
Independent	7.31
GML 23km	1.70
GML 7km	1.69
GML 5km	1.87
GA 23km	2.63
ARTMAP 23km	4.47

The RMS error for the *post priori* probabilities is less than that of the reference regions in each case. These results indicate that, assuming unbiased reference, the accuracy of resulting confusion matrices can be improved by weighting by *post priori* probabilities.

The confusion matrices were then scaled using the scaling coefficients in Table 9-15 for each class.

Table 9-15 Scaling Coefficients for Tank Scene Independent Reference

	GML 23km	GML 7km	GML 5km	GA 23km	ARTMAP 23km
α_1 trail	0.949	1.188	0.934	3.120	0.000
α_2 sand	2.896	2.759	2.967	2.404	1.012
α_3 parking	0.219	0.220	0.218	0.288	0.391
α_4 dirt	0.134	0.130	0.130	0.215	1.905
α_5 roof	0.066	0.066	0.064	0.075	0.096
α_6 road	0.302	0.355	0.342	0.313	0.809
α_7 scrub	1.037	1.025	1.006	0.757	1.120
α_8 forest	1.723	1.745	1.738	1.879	1.480
α_9 driveway	0.535	0.329	0.537	0.217	1.141

The resulting Kappa coefficients of assessments based on the scaled matrices, along with those of the unscaled independent and random sampling are summarized in Table 9-16.

Table 9-16 Kappa Coefficient Accuracy Results of Matrix Scaling Tank Scene

		Random Reference Accuracy	Independent Reference Accuracy	Scaled Independent Reference Accuracy
5km Visibility	GML	0.827	0.983	0.989
7km Visibility	GML	0.827	0.982	0.992
23km Visibility	GML	0.881	0.987	0.992
	Mystic	0.705	0.867	0.841
	ARTMAP	0.651	0.905	0.939

In Figure 9-18, the arithmetic difference between the Kappa coefficients from the scaled independent and random assessments is shown along with the difference between the unscaled independent and random based assessment Kappa values.

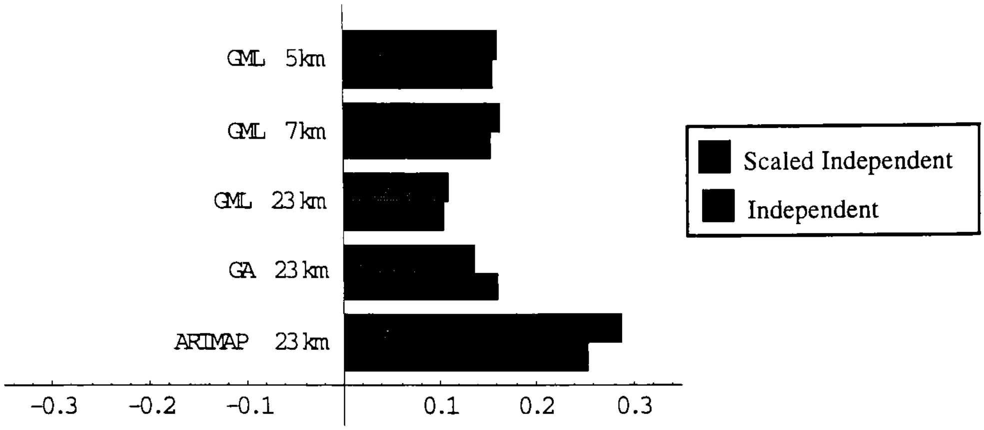

Figure 9-18 Error of Independent and Scaled Independent Assessment of Tank Scene

Again, confusion matrix scaling failed to improve results based on independent reference data. In four out of five cases, scaling actually increased the absolute Kappa error.

9.4.3 Scaling of Desert Scene Confusion Matrices

Nine *desert* scene confusion matrices were scaled by their *post priori* probabilities. The original confusion matrices were based on an independent reference assessment of the scene classified by the GML, Mystic, and Fuzzy ARTMAP classifiers, with GIFOV's of 1m, 2m, and 4 meters.

In Table 9-17, the RMS error between the user selected reference and random sampling marginals and the RMS error between the *post priori* and random sampling probability distributions are shown.

Table 9-17 RMS Error of Marginal Distribution Approximation for Desert Scene

	RMS Error x100		
	1m GIFOV	2m GIFOV	4m GIFOV
Dependent	11.8	17.3	18.4
Independent	7.65	8.99	15.2
GML	2.24	3.39	3.00
GA	4.08	4.44	4.72
ARTMAP	1.31	3.50	5.82

The RMS error for the *post priori* probabilities is less than that of the reference regions in all nine cases. For the *desert* scene, assuming unbiased reference, the accuracy of resulting confusion matrices should be improved by weighting by *post priori* probabilities.

The independent reference confusion matrices were then scaled by the coefficients in Table 9-18.

Table 9-18 Scaling Coefficients for Desert Scene

	1m GIFOV			2m GIFOV			4m GIFOV		
	GML	Mystic	ARTMAP	GML	Mystic	ARTMAP	GML	Mystic	ARTMAP
α_1	1.450	1.517	1.354	1.691	1.585	1.445	2.153	2.503	1.930
α_2	0.651	0.426	0.749	0.196	0.305	0.312	0.649	0.252	0.463
α_3	0.258	0.102	0.146	0.038	0.064	0.054	0.425	0.047	0.063
α_4	0.473	0.038	0.158	0.466	0.031	0.203	0.144	0.011	0.086
α_5	0.213	0.069	1.056	0.368	0.043	1.847	0.905	0.018	1.803
α_6	0.655	1.496	0.239	0.334	1.578	0.361	0.103	0.862	0.078

In Table 9-19 the Tau values for the assessment of all nine images based on random, independent, and scaled independent reference data are summarized.

Table 9-19 Tau Coefficient Results of Matrix Scaling of Desert Scene

		Random Reference	Independent Reference	Scaled Independent Reference
1m GIFOV	GML	0.623	0.950	0.954
	Mystic	0.614	0.796	0.953
	ARTMAP	0.503	0.881	0.932
2m GIFOV	GML	0.577	0.869	0.882
	Mystic	0.573	0.717	0.725
	ARTMAP	0.411	0.870	0.879
4m GIFOV	GML	0.420	0.813	0.889
	Mystic	0.557	0.632	0.582
	ARTMAP	0.218	0.622	0.514

In Figure 9-19, the arithmetic difference between the random and independent reference assessment Tau value is plotted along with the difference between the scaled independent reference.

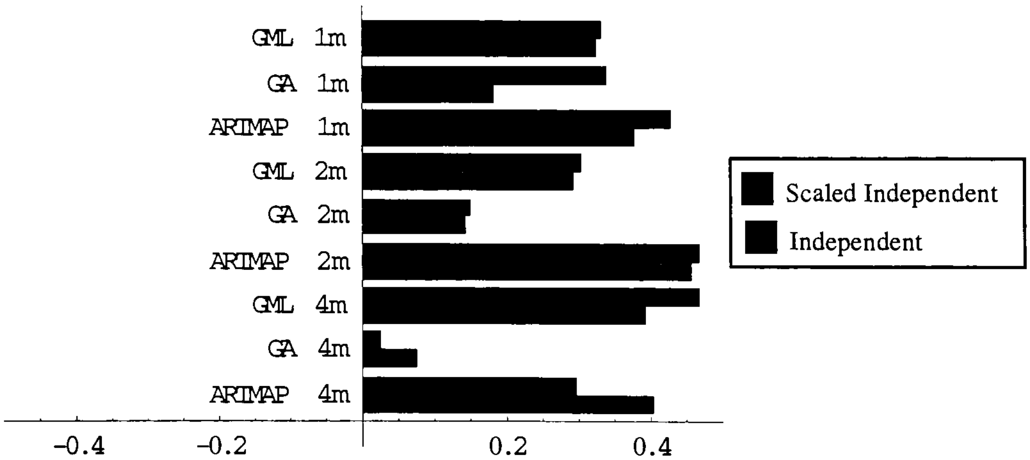

Figure 9-19 Error of Independent and Scaled Independent Assessment of Desert Scene

In all nine cases, both user selected reference sets overestimated the Tau value obtained by random sampling. In only two cases, the Mystic™ and Fuzzy ARTMAP classified 4m GIFOV image, did scaling reduce the Tau error.

9.5 Mystic™ Classifier Performance

9.5.1 Classification Results

The Mystic™ classifier is the newest and least tested of the three classifiers utilized. For this reason, it was of particular interest to this task. On average, Mystic™ proved to be the least accurate of the three classifiers examined while the GML was the most accurate. The average classification accuracy for Mystic™ was only slightly lower than that of the Fuzzy ARTMAP. The Mystic™ classifier was hypersensitive to the size and nature of the training regions used. In fact, early results dictated that training regions for all classifiers be optimized to Mystic™ in a trial and error fashion. However, Mystic™'s sensitivity to noisy training is also coupled with the ability to better handle image data which deviates greatly from a Gaussian spectral probability distribution. This advantage over traditional parametric classifiers, such as the GML, could be realized in situations such as very high resolution imagery. However, this advantage was never realized.

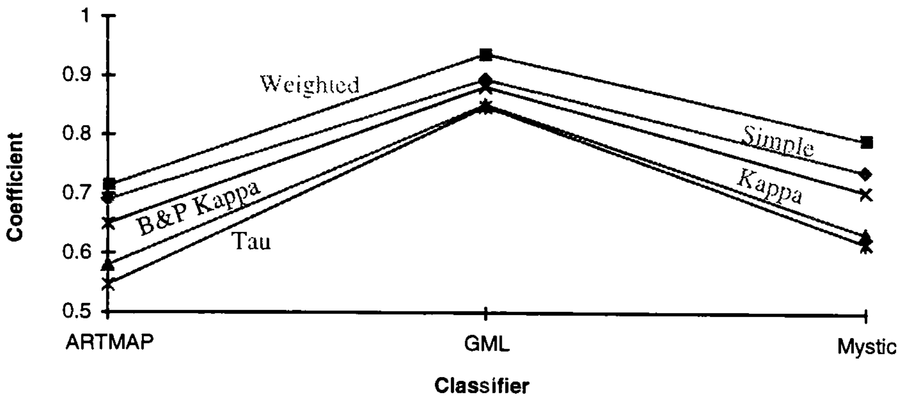

Figure 9-20 Classifier Performance on Desert 23km 1m Image

In Figure 9-20 the results for the *desert* scene with 1 meter resolution, and 23km visibility are shown, based on independent reference data. Figure 9-21 shows the measured accuracy of the *tank* scene generated with randomly sampled reference data.

Mystic™ performance proved to be hit or miss. Experience has shown that Mystic™ accuracy is unpredictable and inconsistent. It was also necessary to try several rules until the appropriate one was identified. Results often showed that each image was best classified by different rules. No one rule could be used on all images and achieve acceptable results. Training the Mystic™ classifier was a painstaking process because two parameters are involved, regions and rules, and the run time was long between trials. In practice, several rules and regions will be used to classify an image and results will be extremely poor. Finally, a combination can often be found which will allow high accuracy. Occasional, accuracies were higher than those of the ARTMAP and GML, for example with the *forest* scene at 23km visibility.

In general, the rules prepackaged with Mystic™ tended to be overly permissive. In other words, the rules for identifying a class often selected too many pixels and failed to reject enough. For this reason, Mystic™ left very few regions undefined. This fact alone slightly reduces measured class map accuracy. Very frequently, more than two rules identified a single pixel as belong to different classes. Mystic™ has a built in system for handling such rule contention. The method of handling can be selected by the user. Usually, the pixel is placed in the class which returned the highest reward factor. The analyst can also select to have the pixels placed in a category according to fixed priorities. Rule contention appears to be a large source of classifier error. During the course of this study, on several occasions, the method of contention handling was changed when class maps contained gross errors. This greatly changed the resulting class map, mainly in the areas of the largest classifier error. Clearly, large numbers of pixels result in conflicting rules. This result indicates that rule contention is positively correlated with locations of classifier error. Mystic™ overall accuracy could likely be increased by reducing the occurrence of conflicting rules. New rules, with fixed upper or lower bounds, could serve this purpose by avoiding the overly permissive nature of the current open ended, thresholding rules.

The Mystic™ classifier also required significantly more time to classify the same images than the other classifiers. This time was needed to optimize the genetic algorithm. Once the optimal rules are found, the time required to apply these rules to an image and generate a class map is relatively short. However, with every new training set, the GA needed to be optimized again. The time required for Mystic™ to classify a scene proved to be very sensitive to the size of the training data supplied. In fact, the training time for Mystic™ grew exponentially with the number of training points supplied, while the training time grew only linearly with the other two classifiers. On one scene in particular, where a large training set was used, Mystic™ required almost two orders of magnitude longer to classify than the GML and Fuzzy ARTMAP.

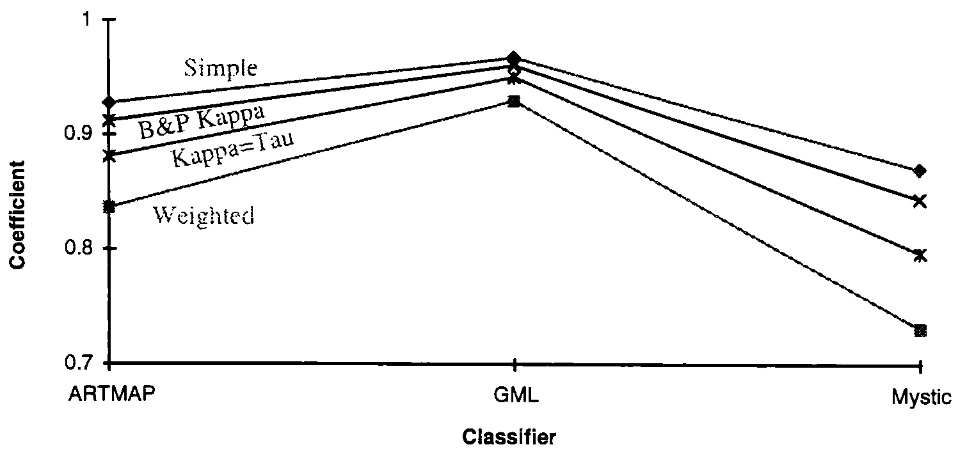

Figure 9-21 Classifier Performance on Tank 23km 1m Image

Mystic™ required not only more computer time but also more analyst time. The GUI for selecting training regions was found to be imprecise and awkward by comparison with contemporary packages such as ENVI™. This problem was compounded by the fact that no provisions where made for importing training data from other sources and the file format specifications were not available for the training region files.

Mystic™ has a limit on the maximum image size from which training data can be collected. The maximum size is 256x256 pixels. The manual states a procedure exists to side-step the limitation by using a process called "compaction". This feature did not function properly and no other method was found to train from larger images. The Matrix™ support personnel were also unable to devise a solution to overcome this limitation. This complicated the importation of training data because most of the baseline images of this study were larger than 256x256 pixels. This problem was solved by training the classifier on a mosaic image of the training regions and then classifying using the optimized rules on the original scene. This additional step was effective but added one more step in an already long process of classification with Mystic™.

The other non-parametric classifier, the Fuzzy ARTMAP, did not fair well compared to the parametric GML classifier. A major source of error for the Fuzzy ARTMAP classifier was its inability to discriminate the metal class category in the *forest* scene under all atmospheric conditions. Even though the classifier was trained with a metal class, it was dropped by the Fuzzy ARTMAP because of its failure to resonate with the proper number of classes. This produced a sequence of zeros in one of the five rows of the resulting confusion matrices.

9.5.2 Suggestions for Improvement

Mystic™ results indicate that the classification algorithm has promising potential but is not yet optimized for use on an operational basis. For example, the extended training time required by Mystic™ is not offset by equally high classification. It is also apparent that an overall gain in Mystic™ accuracy could be attained by expanding the scope of the GA to also optimize which rule is selected. Currently, a user must guess the best rule by experience. An inexperienced user has no knowledge about how to select the best rule so randomly selecting rules and training with each is the only technique available. This very optimization procedure is exactly what GA specialize in.

In its present state Mystic™ is a simple "per pixel" classifier. However, it could be expanded to use other contextual information such as texture, as could the GML and Fuzzy ARTMAP. Over the years, several texture based metrics have been presented in the remote sensing and pattern recognition literature (Haralick *et al*, 1973; Galloway, 1975; Wang, 1990; Stormberg and Farr, 1986). In the case of Mystic™, rules could be generated that would use statistical measures such as the range or variance over a local (windowed) area to also classify on the basis of texture. To avoid correlation, often encountered in the statistical classification of multispectral or hyperspectral imagery, the texture rules could be used only on the first band of the Principal Component transformation of the entire image. This procedure would reduce redundant information and significantly decrease the run time because the statistical rule would only need to run on one band. Hybrid rules could also make decisions on the basis of spectral and spatial content. Swain (1978) and Rosenblum (1990) have presented models for optimizing the balance between classification based on textural and spectral information. However, rule based classifiers are limited in the number of parameters that can be classified and remain tractable in a reasonable time period. Adding parameters can increase the size of the search space and the time required to find the solution geometrically. Heuristics could be introduced to reduce this search space.

The design philosophy behind Mystic™ suggests that once optimal rules are found on one training image they can be applied to other images. This would make the classification of subsequent images faster than the two competing algorithms because the rules would only need to be optimized once. This procedure is contingent on the utilization of imagery that is highly similar (atmosphere, gain, bias, content, view angle) to the training image. The accuracy of this method is questionable even with substantial preprocessing such as atmospheric correction to obtain images in reflectance units.

Experience with Mystic™ has indicated, like other non-parametric classifiers, that it is hypersensitive to noisy training data. For this reason, small training sets proved to be most effective. Mystic™'s high sensitivity to stressing situations, such as atmosphere, further implies that training on images utilizing training data collected from another image will be error prone. While it was not explicitly examined as part of this task, results indicate that using training data collected from another image would be problematic.

The incorporation of an 'OR' logical operation into the rules would allow a classifier which wouldn't need to split spectrally bimodal classes. Other possible improvements include more rules to select from, including more exclusive rules, combination or hybrid rules, and implementing a hybrid classification technique utilizing Mystic™ in addition to a more traditional parametric classifier such as the GML. Such a classifier could benefit from the advantages of both the rule based classifier and the statistical classifier.

10. Summary & Conclusion

The goal of this research was to quantify the exploitation techniques of multispectral classification and its accuracy assessment. In the attempt to characterize the classifiers, accuracy assessment techniques, and the effect of stressing parameters, several critical issues were raised. They ranged from the merit of the classification accuracy metrics, the robustness of the input imagery data sets, and the statistical significance of the conclusions and trends drawn from sparse set of scenarios.

User selected reference data was shown to consistently overestimate the true accuracy. Dependent data was slightly more overly optimistic than independent data, but both were significantly worse than random sampling reference data. The effect of spatial resolution on classification accuracy was the opposite of the expected result. Classification accuracy increased as the spatial resolution of the imagery increased. This results was due to the spectral pixel mixing involved in degrading the very high resolution aerial line scanner imagery. The effects of atmospherics verified that there is a decrease in classification accuracy as visibility decreased. The overall rank of performance based on the applied classification metrics places GML as the most accurate followed by Fuzzy ARTMAP, and then by the Mystic™ classifier. It should be noted, however, that many of these findings are drawn from relatively few treatments.

In terms of the classification metrics, the general rank from highest accuracy to lowest were Simple accuracy, Brennan and Prediger's Kappa, Kappa coefficient, and the Tau coefficient. Historically, because of the numerous methods of accuracy assessment it has been difficult to compare the accuracy of data sets used in different research papers. This has been a common complaint because of clear examples where large discrepancies in accuracy were reported in different articles while utilizing the same data set. One solution for this has been to report several different accuracy metrics for each class map. However, many studies still report only one metric, fail to report the error matrix, and do not acknowledge the potential for bias in the sampling method (Story, 1986). Recently, the Kappa coefficient with reference data collected from stratified random sampling has become the most popular combination for assessing accuracy. Both the Kappa and stratified random sampling have been shown to provide accurate results with low bias. While this sampling method and accuracy metric have recently emerged as common approach they have not nearly obtained the status of an industry wide standard. In addition, in this research the Kappa coefficient was shown to be a misleading metric for representing class map accuracy. Increases in class map accuracy can actually result in lower Kappa values.

What this task has shown is that the current set of classification metrics based on the confusion matrix provides only a limited reflection of the classification performance. Overall, the matrices tracked each other in terms of the measured accuracy. Each included some factor such as the probability of correct classification due to chance or placed some assumptions on *a priori* probabilities. But, in general, the trends were the same. The choice of which metric to use is dependent on which prerequisite conditions can be adequately fulfilled by the selected method of reference data collection. These variations on the Simple confusion matrix accuracy mainly provide a global measure of the classification performance with minimal regard to the spatial structure of the classification. This is understandable in light of the fact that a majority of the applications using these metrics concern themselves with large scale, low spatial resolution overhead imagery. In these scenarios, coarse classifications are generally sufficient to produce a land cover class map for purposes of agricultural and urban planning. At such scales, any changes that are perceived as real and significant are typically gross in size and have roots in proportionately gross causes (e.g. conversion of agricultural fields due to urban development). These metrics have been adequately applied for these cases because the penalties for misclassification have not been severe, thus tolerating significant margins of error. When these metrics are applied to high spatial resolution data sets, however, where the exploitation task often involves the detection of a small concealed target against a background expanse, the need for an accurate classification assessment becomes extremely critical. Detection avoidance being the goal of the concealed target, misclassified pixels will require additional scrutiny to rank the misclassifications and provide an indicator of target probability. In these cases, the location information becomes a critical factor because the spatial context in which a pixel is misclassified provides important clues as to the factors producing the misclassifications.

Confusion matrix marginal distribution scaling is useful in cases where disproportionate reference sampling is used along with accuracy metrics which depend on proper marginal distributions. User selected and stratified random sampling both yield disproportionate class distributions. The Simple percentage, Kappa, Brennan and Prediger's Kappa, Tau, and User's accuracy coefficients are examples of accuracy metrics who accuracy depends on proportionate sampling. Properly sampled reference data, such as a pure random sampling, does not benefit from confusion matrix scaling. Error metrics such as the Weighted or Producer's accuracy coefficients which do not depend on the marginal distribution of the reference data and therefore also fail to benefit from matrix scaling. For the three scenes on which it was attempted, confusion matrix scaling by *post priori* probabilities did not reliably increase the accuracy of the assessment procedure. By chance, in each case, scaling accentuated the contribution of a more biased than average class. This result was due to bias inherent in user selected reference, not improper probability distributions. The normalized histogram probabilities were shown to accurately estimate the true *a priori* distribution using Kolmogorov-Smirnov two-sample test and RMS error as a measure. In addition, the user selected reference marginals were shown to consistently serve as a poor estimator of the *a priori* distribution. Estimates of the true class distribution for the purpose of confusion matrix distribution scaling could come from other sources as well. For example, previous studies, small random samplings, or even analyst estimates.

One of the difficulties identified is the lack of a robust set of real image data with sufficient diversity to adequately perform a sensitivity analysis. While this is may be addressed by tasking for additional data collections, the resource demands would make such an effort logistically and fiscally prohibitive. Such a data set would, in all practicality, contain variabilities in sensor geometry, atmospheric conditions, etc, that would be sources of uncertainty. Despite all the various metrics investigated in this report, it became evident early in the research that there did not exist a standard or baseline from which an absolute measure can be computed. While field measurements were taken by other investigators for verification, these still consisted of discrete samplings prone to levels of statistical uncertainties. In an attempt to emulate truth and eliminate (or at least quantify) these uncertainties, a series of DIRSIG simulated images were used as input into the classifiers as baseline cases. The ability to create a suite of tightly controlled imagery for use as input into the classifiers offers a potentially valuable method to test these classification algorithms. Not only is "truth" inherently known, but the capability of creating various scene types (urban, forest, etc.) and to do so for a near infinite number of imaging variations (time of day, atmosphere, etc.) provides an opportunity to definitively address both the classifiers and the classification metrics.

11. Recommendations for Future Work

- Evaluate accuracy of stratified random sampling with confusion matrix scaling
- Implement Suggestions for Mystic™ or other rule based genetic algorithm classifier
- Quantify the increase in classification accuracy using image fusion as preprocessor
- Using more treatments and ANOVA approach, determine significance of stressing parameters
- Further use of synthetic imagery to allow quantitative assessment of classifiers and stressing parameters
- Utilize test imagery with real rather than simulated stressing parameters
- Automate training procedure using clustering algorithm to remove effect of user from analysis of spatial resolution
- Analyze effect of spatial resolution on classification accuracy based on variety of original image resolutions from high altitude aerial sensor to satellite imagery

12. References

Agresti, A. *Categorical Data Analysis* John Wiley & Sons Inc., New York 1990

Allen, J.R., Raqueño, R., and Schott, J.R. "Terrain Categorization Error Analysis" Rochester Institute of Technology, Feb 1997, DIRS Report #96/97-72-154

Aronoff, S. "The Minimum Accuracy Values as an Index of Classification Accuracy" Photogrammetric Engineering and Remote Sensing Vol.51, #1, Jan 1985 99-111

Avery, T.E. and Berlin, G.L. *Interpretation of Aerial Photographs* Fourth Edition, Burgess, Minneapolis 1985

Bauer, M.E., Burk, T.E., Ek, A.R., Coppin, P.R., Lime, S.D., Walsh, T.A., Walters, D.K., Beford, W., and Heinzen, D.F. "Satellite Inventory of Minnesota Forest Resources" Photogrammetric Engineering and Remote Sensing, Vol.60 #3 Mar 1994 287-298

Birnbaum, Z.W. and Hall, R.A. "Small-Sample Distribution for Multi-Sample Statistics of the Smirnov Type" Ann. Math. Statist. #31 1960 710-720

Brennan, R.L. and Prediger, D.J. "Coefficient Kappa: Some Uses, Misuses, and Alternatives" Educational and Psychological Measurement #41 1981 687-699

Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., and Rosen, D.B. "Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps" IEEE Transactions on Neural Networks Vol. 3 #5 Sept 1992

Cohen, J. "A Coefficient of Agreement for Nominal Scales " Educational and Psychological Measurement #20 1960 37-46

Conese, C. and Maselli F. "Use of Error Matrices to Improve Area Estimates with Maximum Likelihood Classification Procedures" Remote Sensing of Environment Vol.40 #2 May 1992 113-124

Congalton, R.G. "A Comparison of Sampling Schemes Used in Generating Error Matrices for Assessing the Accuracy of Maps Generated from Remotely Sensed Data" Photogrammetric Engineering and Remote Sensing Vol.54 #5 May 1988 593-600

Congalton, R.G. and Mead, R.A. "A Quantitative Method to Test for Consistency and Correctness in Photointerpretation" Photogrammetric Engineering and Remote Sensing Vol.49 #1 Jan 1983 69-74

Congalton, R.G., Mead, R.A., and Oderwald, R. "Statistical Techniques for the Analysis of Landsat Classification Accuracy" Presented at the ASPRS/ACSM Annual Meeting, St. Louis, Missouri 1980

Congalton, R.G., Oderwald, R.D., and Mead, R.A. "Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques" Photogrammetric Engineering and Remote Sensing Vol.49 #12,Dec 1983 1671-1678

Duda, R.O. and Hart, P.E. *Pattern Classification and Scene Analysis* John Wiley & Sons, New York 1973

Fitzpatrick-Lins, K. "The Accuracy of Selected Land Use and Land Cover Maps at Scales f 1:250,000 and 1:100,000" Geological Survey Circular 829 1980

Foody, G.M. "On the Compensation for Chance Agreement in Image Classification Accuracy Assessment" Photogrammetric Engineering and Remote Sensing Vol.58 #10 Oct 1992 1459-1460

Franklin, S.E. and Wilson, B.A. "A Three-Stage Classifier for Remote Sensing of Mountain Environments" Photogrammetric Engineering and Remote Sensing Vol.58 #4 Apr 1992 449-454

Friedberg, R.M. "A Learning Machine: Part I" IBM Journal, Vol.2 1958 2-12

Galloway, M.M. "Texture Classification Using Gray Level Run Lengths" Computer Graphics and Image Processing Vol.4 1975 172-79

Gong, P. and Howarth, P.J. "An Assessment of Some Factors Influencing Multispectral Land-Cover Classification" Photogrammetric Engineering and Remote Sensing Vol.56 #5 May 1990 597-603

Hammond, T.O. and Verbyla, D.L. "Optimistic Bias in Classification Accuracy Assessment" International Journal of Remote Sensing Vol.17 #6 Apr 1996 1261-1266

Haralick, R.M., Shanmugam, K., and Dinstein, I. "Textural Features of Image Classification" IEEE Transactions on Systems, Man, and Cybernetics Vol.6 SMC-3 1973

Hord, R.M., and Brooner, W. "Land-Use Map Accuracy Criteria" Photogrammetric Engineering and Remote Sensing Vol.42 #5 May 1976 671-677

Hudson, W.D. and Ramm, C.W. "Correct Formulation of the Kappa Coefficient of Agreement" Photogrammetric Engineering and Remote Sensing Vol.53 #4 Apr 1987 421-422

Irons, J.R., Markham, B.L., Nelson, R.F., Toll, D.L., Williams, D.L., Latty, R.S., and Stauffer, M.L. "The Effects of Spatial Resolution on the Classification of Thematic Mapper Data" International Journal of Remote Sensing Vol.6 #8 1985 1385-1403

Janssen, L.L.F., Van Der Wel, J.M. "Accuracy Assessment of Satellite Derived Land-Cover Data: A Review" Photogrammetric Engineering and Remote Sensing Vol.60 #4 Apr 1994 419-426

Klecka, W.R. *Discriminant Analysis* SAGE Publication, Inc., Beverly Hills, CA 1980

Ma, Z. and Redmond, R.L. "Tau Coefficients for Accuracy Assessment of Classification of Remote Sensing Data" Photogrammetric Engineering and Remote Sensing Vol.61 #4 Apr 1995 435-439

Marsh, S.E., Walsh, J.L., and Sobrevila, C. "Evaluation of Airborne Video Data for Land-Cover Classification Accuracy Assessment in an Isolated Brazilian Forest" Remote Sensing of Environment Vol.48 #1 Apr 1994 61-69

Nageswara Rao, P.P. and Mohankumar, A. "Cropland Inventory in the Command Area of Krishnarajasagar Project Using Satellite Data" International Journal of Remote Sensing Vol.15 #6 1994 1295-1305

Nessmiller, S.W. "A Comparison of the Performance of Non-Parametric Classifiers with Gaussian Maximum Likeihood for the Classification of Multispectral Remotely Sensed Data" Masters Thesis, Rochester Institute of Technology, Center for Imaging Science, Sept 1995

Raqueño, R., Konno, D.X., and Farnung, C.E. "Influence of MTF on Exploitation Accuracy" Feb 1997 RIT/DIRS Report #96/97-63-155

Rosenblum, W. "Optimal Selection of Textural and Spectral Features for Scene Segmentation" Masters Thesis, Center for Imaging Science, Rochester Institute of Technology, Rochester, New York 1990

Rosenfield, G.H. and Fitzpatrick-Lins, K. "A Coefficient of Agreement as a Measure of Thematic Classification Accuracy" Photogrammetric Engineering and Remote Sensing Vol. 58, #2, Feb 1986 223-227

Rosenfield, G.H. and Melly, M.L. "Application of Statistics to Thematic Mapping" Photogrammetric Engineering and Remote Sensing Vol.48 #5 1980 1287-1294

Rutchey, K and Vilcheck, L. "Development of an Everglades Vegetation Map Using a SPOT Image and the Global Positioning System" Photogrammetric Engineering and Remote Sensing Vol.60 #6 Jun 1994 767-775

Senseman, G.M., Bagley, C.F., and Tweddale, S.A. "Accuracy Assessment of the Discrete Classification of Remotely-Sensed Digital Data for landcover Mapping" USACERL Technical Report EN-95/04

Schott, J.R. *Remote Sensing: The Imaging Chain Approach* Oxford University Press 1997

Schott, J.R., Mason, J.E., Salvaggio, C., Sirianni, J.D., Rose, R.A., Kulp, E.O., and Rankan, D.K. "DIRSIG, Digital Imaging and Remote Sensing Image Generation, Description, Enhancement, and Validation" Jul 1993 DIRS Report #92-93-51-146

Stromberg, W.D. and Farr, T.G. "A Fourier-Based Textural Feature Extraction Procedure" IEEE Transactions on Geoscience and Remote Sensing Vol. GE-24 #5 1986 722-31

Story, M. "Accuracy Assessment: A User's Perspective" Photogrammetric Engineering and Remote Sensing Vol.52 #3 Mar 1986 397-399

Swain, P.H. *Fundamentals of Pattern Recognition in Remote Sensing* In P.H. Swain and S.M. Davis, eds. *Remote Sensing: The Quantitative Approach*, McGraw-Hill, New York 1978

Szajgin, J., Pettinger, L.R., Linden, D.S., and Ohlen, D.O. "Arizona Vegetation Resource Inventory (AVRI) Accuracy Assessment: Final Report" Geological Survey Open File Report No.82-814 1982

Thomasson, J.A., Bennett, C.W., Jackson, B.D., and Mailander, M.P. "Differentiating Bottomland Tree Species with Multispectral Videography" Photogrammetric Engineering and Remote Sensing Vol.60 #1 Jan 1994 55-59

Tou, J.T., and Gonzalez, R.C. *Pattern Recognition Principals* Addison-Wesley Publishing Company, Reading, Massachusetts 1974

Van Genderen, J.L., Lock, B.F., and Vass, P.A. "Remote Sensing: Statistical Testing of Thematic Map Accuracy" Remote Sensing of Environment Vol. 7 1978 3-14

Verbyla, D.L. and Hammond, T.O. "Conservative Bias in Classification Accuracy Assessment Due to Pixel-by-Pixel Comparison of Classified Images with Reference Grids" International Journal of Remote Sensing Vol.16 #3 Feb 1995 581-587

Wang, L, and He, D. "A New Statistical Approach for Textural Analysis" Photogrammetric Engineering and Remote Sensing Vol.56 #1 1990

Wayne, W.D. *Applied Nonparametric Statistics* PWS-KENT Publishing Company, Boston 1990

13. Appendix A

Scene	Resolution	ATM	Classifier	Reference	Page
Tank	1m	5km	GML	Dependent	154
Tank	1m	7km	GML	Dependent	155
Tank	1m	23km	GA	Dependent	156
Tank	1m	23km	GML	Dependent	157
Tank	1m	23km	ARTMAP	Dependent	158
Tank	1m	5km	GML	Independent	159
Tank	1m	7km	GML	Independent	160
Tank	1m	23km	GML	Independent	161
Tank	1m	23km	GA	Independent	162
Tank	1m	23km	ARTMAP	Independent	163
Tank	1m	5km	GML	Scaled Independent	164
Tank	1m	7km	GML	Scaled Independent	165
Tank	1m	23km	GML	Scaled Independent	166
Tank	1m	23km	GA	Scaled Independent	167
Tank	1m	23km	ARTMAP	Scaled Independent	168
Tank	1m	5km	GML	Random	169
Tank	1m	7km	GML	Random	170
Tank	1m	23km	GML	Random	171
Tank	1m	23km	GA	Random	172
Tank	1m	23km	ARTMAP	Random	173
Forest	1m	5km	GML	Dependent	174
Forest	1m	5km	GA	Dependent	175
Forest	1m	5km	ARTMAP	Dependent	176
Forest	1m	7km	GML	Dependent	177
Forest	1m	7km	GA	Dependent	178
Forest	1m	7km	ARTMAP	Dependent	179
Forest	1m	23km	GML	Dependent	180
Forest	1m	23km	GA	Dependent	181
Forest	1m	23km	ARTMAP	Dependent	182
Forest	1m	5km	GML	Independent	183
Forest	1m	5km	GA	Independent	184
Forest	1m	5km	ARTMAP	Independent	185
Forest	1m	7km	GML	Independent	186
Forest	1m	7km	GA	Independent	187
Forest	1m	7km	ARTMAP	Independent	188
Forest	1m	23km	GML	Independent	189
Forest	1m	23km	GA	Independent	190
Forest	1m	23km	ARTMAP	Independent	191
Forest	1m	5km	GML	Scaled Independent	192
Forest	1m	5km	GA	Scaled Independent	193
Forest	1m	5km	ARTMAP	Scaled Independent	194
Forest	1m	7km	GML	Scaled Independent	195
Forest	1m	7km	GA	Scaled Independent	196
Forest	1m	7km	ARTMAP	Scaled Independent	197
Forest	1m	23km	GML	Scaled Independent	198
Forest	1m	23km	GA	Scaled Independent	199
Forest	1m	23km	ARTMAP	Scaled Independent	200

Scene	Resolution	ATM	Classifier	Reference	Page
Forest	1m	5km	GML	Random	201
Forest	1m	5km	GA	Random	202
Forest	1m	5km	ARTMAP	Random	203
Forest	1m	7km	GML	Random	204
Forest	1m	7km	GA	Random	205
Forest	1m	7km	ARTMAP	Random	206
Forest	1m	23km	GML	Random	207
Forest	1m	23km	GA	Random	208
Forest	1m	23km	ARTMAP	Random	209
Forest	1m	5km	GML	Synthetic	210
Forest	1m	5km	GA	Synthetic	211
Forest	1m	5km	ARTMAP	Synthetic	212
Forest	1m	7km	GML	Synthetic	213
Forest	1m	7km	GA	Synthetic	214
Forest	1m	7km	ARTMAP	Synthetic	215
Forest	1m	23km	GML	Synthetic	216
Forest	1m	23km	GA	Synthetic	217
Forest	1m	23km	ARTMAP	Synthetic	218
Forest	1m	5km	GML	Dependent/Synthetic	219
Forest	1m	5km	GA	Dependent/Synthetic	220
Forest	1m	5km	ARTMAP	Dependent/Synthetic	221
Forest	1m	7km	GML	Dependent/Synthetic	222
Forest	1m	7km	GA	Dependent/Synthetic	223
Forest	1m	7km	ARTMAP	Dependent/Synthetic	224
Forest	1m	23km	GML	Dependent/Synthetic	225
Forest	1m	23km	GA	Dependent/Synthetic	226
Forest	1m	23km	ARTMAP	Dependent/Synthetic	227
Forest	1m	5km	GML	Independent/Synthetic	228
Forest	1m	5km	GA	Independent/Synthetic	229
Forest	1m	5km	ARTMAP	Independent/Synthetic	230
Forest	1m	7km	GML	Independent/Synthetic	231
Forest	1m	7km	GA	Independent/Synthetic	232
Forest	1m	7km	ARTMAP	Independent/Synthetic	233
Forest	1m	23km	GML	Independent/Synthetic	234
Forest	1m	23km	GA	Independent/Synthetic	235
Forest	1m	23km	ARTMAP	Independent/Synthetic	236
Desert	1m	23km	GML	Dependent	237
Desert	1m	23km	GA	Dependent	238
Desert	1m	23km	ARTMAP	Dependent	239
Desert	2m	23km	GML	Dependent	240
Desert	2m	23km	GA	Dependent	241
Desert	2m	23km	ARTMAP	Dependent	242
Desert	4m	23km	GML	Dependent	243
Desert	4m	23km	GA	Dependent	244
Desert	4m	23km	ARTMAP	Dependent	245

Scene	Resolution	ATM	Classifier	Reference	Page
Desert	1m	23km	GML	Independent	246
Desert	1m	23km	GA	Independent	247
Desert	1m	23km	ARTMAP	Independent	248
Desert	2m	23km	GML	Independent	249
Desert	2m	23km	GA	Independent	250
Desert	2m	23km	ARTMAP	Independent	251
Desert	4m	23km	GML	Independent	252
Desert	4m	23km	GA	Independent	253
Desert	4m	23km	ARTMAP	Independent	254
Desert	1m	23km	GML	Scaled Independent	255
Desert	1m	23km	GA	Scaled Independent	256
Desert	1m	23km	ARTMAP	Scaled Independent	257
Desert	2m	23km	GML	Scaled Independent	258
Desert	2m	23km	GA	Scaled Independent	259
Desert	2m	23km	ARTMAP	Scaled Independent	260
Desert	4m	23km	GML	Scaled Independent	261
Desert	4m	23km	GA	Scaled Independent	262
Desert	4m	23km	ARTMAP	Scaled Independent	263
Desert	1m	23km	GML	Random	264
Desert	1m	23km	GA	Random	265
Desert	1m	23km	ARTMAP	Random	266
Desert	2m	23km	GML	Random	267
Desert	2m	23km	GA	Random	268
Desert	2m	23km	ARTMAP	Random	269
Desert	4m	23km	GML	Random	270
Desert	4m	23km	GA	Random	271
Desert	4m	23km	ARTMAP	Random	272

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	248095
Pixels Classified :	232241
Percent Classified :	93.6

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	193	0	0	0	0	0	0	0	0	193
sand	0	246	0	0	0	0	0	0	0	246
parking	0	0	341	0	0	0	0	0	0	341
dirt	0	0	0	403	0	0	0	0	0	403
roof	0	0	0	0	244	0	0	0	0	244
road	0	0	0	0	0	335	0	0	2	337
scrub	0	0	0	0	0	0	720	3	0	723
forest	0	0	0	0	0	0	0	376	0	376
driveway	0	0	0	0	0	2	0	0	66	68
	193	246	341	403	244	337	720	379	68	

Reference Source / # :	Dependent - 2948	
Pixels Verified :	2931	
Diagonal Elements :	2924	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.998	(0.996 , 0.999)
Weighted Accuracy (w) :	0.995	(0.991 , 1.000)
Kappa ($\hat{k}$) :	0.997	(0.995 , 0.999)
B&P's Kappa ($\hat{k}_n$) :	0.997	(0.995 , 0.999)
Tau (T_p) :	0.997	(0.995 , 0.999)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	248095
Pixels Classified :	228878
Percent Classified :	92.3

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	193	0	0	0	0	0	0	0	0	193
sand	0	246	0	0	0	0	0	0	0	246
parking	0	0	342	0	0	0	0	0	0	342
dirt	0	0	0	403	0	0	0	0	0	403
roof	0	0	0	0	260	0	0	0	0	260
road	0	0	0	0	0	337	0	0	0	337
scrub	0	0	0	0	0	0	720	0	0	720
forest	0	0	0	0	0	0	0	379	0	379
driveway	0	0	0	0	0	0	0	0	68	68
	193	246	342	403	260	337	720	379	68	

Reference Source / # :	Dependent - 2948
Pixels Verified :	2948
Diagonal Elements :	2949

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	1.0	(1.0 , 1.0)
Weighted Accuracy (w) :	1.0	(1.0 , 1.0)
Kappa ($\hat{k}$) :	1.0	(1.0 , 1.0)
B&P's Kappa ($\hat{k}_n$) :	1.0	(1.0 , 1.0)
Tau (T_p) :	1.0	(1.0 , 1.0)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold 2 bands rule with large training set
Total Pixels :	248095
Pixels Classified :	262144
Percent Classified :	105.7

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	178	5	0	3	0	0	0	0	0	186
sand	0	241	1	0	2	0	0	0	0	244
parking	0	0	334	0	11	0	0	0	0	345
dirt	15	0	0	400	0	0	0	0	0	415
roof	0	0	2	0	247	0	0	0	0	249
road	0	0	0	0	0	337	0	0	60	397
scrub	0	0	5	0	0	0	716	1	1	723
forest	0	0	0	0	0	0	4	378	0	382
driveway	0	0	0	0	0	0	0	0	7	7
	193	246	342	403	260	337	720	379	68	

Reference Source / # :	Dependent - 2948
Pixels Verified :	2948
Diagonal Elements :	2838

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.963	(0.956 , 0.970)
Weighted Accuracy (w) :	0.880	(0.870 , 0.890)
Kappa ($\hat{k}$) :	0.956	(0.948 , 0.964)
B&P's Kappa ($\hat{k}_n$) :	0.958	(0.950 , 0.966)
Tau (T_p) :	0.957	(0.949 , 0.965)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	248095
Pixels Classified :	222617
Percent Classified :	89.7

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	193	0	0	0	0	0	0	0	0	193
sand	0	246	0	0	0	0	0	0	0	246
parking	0	0	340	0	0	0	0	0	0	340
dirt	0	0	0	403	0	0	0	0	0	403
roof	0	0	0	0	241	0	0	0	0	241
road	0	0	0	0	0	334	0	0	2	336
scrub	0	0	0	0	0	0	720	4	0	724
forest	0	0	0	0	0	0	0	375	0	375
driveway	0	0	0	0	0	3	0	0	66	69
	193	246	340	403	241	337	720	379	68	

Reference Source / # :	Dependent - 2948
Pixels Verified :	2927
Diagonal Elements :	2918

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.997	(0.995 , 0.999)
Weighted Accuracy (w) :	0.995	(0.990 , 0.999)
Kappa ($\hat{k}$) :	0.996	(0.994 , 0.999)
B&P's Kappa ($\hat{k}_n$) :	0.997	(0.994 , 0.999)
Tau (T_p) :	0.996	(0.994 , 0.999)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	248095
Pixels Classified :	247396
Percent Classified :	99.7

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	0	0	0	0	0	0	0	0	0	0
sand	0	246	0	0	0	0	0	0	0	246
parking	0	0	342	0	0	0	0	0	0	342
dirt	0	0	0	402	0	0	0	0	0	402
roof	0	0	0	0	260	0	0	0	0	260
road	0	0	0	0	0	292	0	0	0	292
scrub	0	0	0	0	0	0	718	0	0	718
forest	193	0	0	1	0	0	2	379	0	575
driveway	0	0	0	0	0	45	0	0	68	113
	193	246	342	403	260	337	720	379	68	

Reference Source / # :	Dependent - 2948	
Pixels Verified :	2948	
Diagonal Elements :	2707	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.918	(0.908 , 0.928)
Weighted Accuracy (w) :	0.873	(0.869 , 0.878)
Kappa ($\hat{k}$) :	0.905	(0.893 , 0.916)
B&P's Kappa ($\hat{k}_n$) :	0.908	(0.897 , 0.919)
Tau (T_p) :	0.905	(0.893 , 0.916)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	248095
Pixels Classified :	232241
Percent Classified :	0.936

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	268	0	0	44	0	0	0	0	0	312
sand	0	507	0	0	0	0	0	0	0	507
parking	0	0	913	0	0	0	0	0	0	913
dirt	0	0	0	309	0	0	0	0	0	309
roof	0	0	0	0	669	0	0	0	0	669
road	0	0	0	0	0	757	0	0	25	782
scrub	0	0	0	0	0	3	2032	0	0	2035
forest	0	0	0	0	0	0	23	1673	0	1696
driveway	0	0	2	0	0	4	0	0	135	141
	268	507	915	353	669	764	2055	1673	160	

Reference Source / # :	Independent - 7434	
Pixels Verified :	7364	
Diagonal Elements :	7263	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.986	(0.984 , 0.989)
Weighted Accuracy (w) :	0.966	(0.959 , 0.974)
Kappa ($\hat{k}$) :	0.983	(0.980 , 0.987)
B&P's Kappa ($\hat{k}_n$) :	0.985	(0.982 , 0.988)
Tau (T_p) :	0.983	(0.980 , 0.987)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	248095
Pixels Classified :	228878
Percent Classified :	92.3

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	268	0	0	52	0	0	0	0	0	320
sand	0	507	0	0	0	0	0	0	0	507
parking	0	0	913	0	0	0	0	0	0	913
dirt	0	0	0	302	0	0	0	0	0	302
roof	0	0	0	0	665	0	0	0	0	665
road	0	0	0	0	0	760	0	0	32	792
scrub	0	0	0	0	0	2	2031	0	0	2033
forest	0	0	0	0	0	0	23	1674	0	1697
driveway	0	0	0	0	0	1	0	0	128	129
	268	507	913	354	665	763	2054	1674	160	

Reference Source / # :	Independent - 7434	
Pixels Verified :	7358	
Diagonal Elements :	7248	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.985	(0.982 , 0.988)
Weighted Accuracy (w) :	0.960	(0.952 , 0.968)
Kappa ($\hat{k}$) :	0.982	(0.979 , 0.985)
B&P's Kappa ($\hat{k}_n$) :	0.983	(0.980 , 0.986)
Tau (T_p) :	0.982	(0.979 , 0.985)

Scene Name : Tank Scene

Image Dimensions : 512 x 512 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Gaussian Maximum Likelihood

Notes : 0.15 Probability Threshold

Total Pixels : 248095

Pixels Classified : 222617

Percent Classified : 89.7

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	268	0	0	29	0	0	0	0	0	297
sand	0	507	0	0	0	0	0	0	0	507
parking	0	0	913	0	0	0	0	0	0	913
dirt	0	0	0	319	0	0	0	0	0	319
roof	0	0	0	0	653	0	0	0	0	653
road	0	0	0	0	0	757	0	0	20	777
scrub	0	0	0	0	0	2	2029	0	0	2031
forest	0	0	0	0	0	0	22	1673	0	1695
driveway	0	0	1	0	0	4	0	0	140	145
	268	507	914	348	653	763	2051	1673	160	

Reference Source / # : Independent - 7434

Pixels Verified : 7337

Diagonal Elements : 7259

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.989	(0.987 , 0.992)
Weighted Accuracy (w) :	0.975	(0.968 , 0.981)
Kappa ($\hat{k}$) :	0.987	(0.984 , 0.990)
B&P's Kappa ($\hat{k}_n$) :	0.988	(0.985 , 0.991)
Tau (T_p) :	0.987	(0.984 , 0.990)

Scene Name : Tank Scene

Image Dimensions : 512 x 512 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Rule Based Genetic Algorithm (MYSTIC)

Notes : threshold 2 bands rule with large training set

Total Pixels : 248095

Pixels Classified : 262144

Percent Classified : 105.7

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	255	210	0	72	0	0	0	0	0	537
sand	0	296	10	3	4	5	124	40	0	482
parking	0	1	898	0	18	4	0	0	0	921
dirt	13	0	0	281	0	0	0	0	0	294
roof	0	0	0	0	680	0	0	0	0	680
road	0	0	0	0	0	727	0	0	131	858
scrub	0	0	7	0	1	53	1832	0	22	1915
forest	0	0	0	0	0	2	104	1634	0	1740
driveway	0	0	0	0	0	0	0	0	7	7
	268	507	915	356	703	791	2060	1674	160	

Reference Source / # : Independent - 7434

Pixels Verified : 7434

Diagonal Elements : 6610

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.889	(0.882 , 0.896)
Weighted Accuracy (w) :	0.789	(0.780 , 0.798)
Kappa ($\hat{k}$) :	0.867	(0.858 , 0.875)
B&P's Kappa ($\hat{k}_n$) :	0.875	(0.867 , 0.883)
Tau (T_p) :	0.866	(0.858 , 0.875)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	248095
Pixels Classified :	247396
Percent Classified :	99.7

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	0	0	0	0	0	0	0	0	0	0
sand	0	503	0	0	0	0	0	0	0	503
parking	0	0	912	0	5	10	0	0	2	929
dirt	11	1	0	314	0	0	0	0	0	326
roof	0	0	3	0	463	0	0	0	0	466
road	0	0	0	0	0	673	0	44	0	717
scrub	0	0	0	0	0	16	1994	7	0	2017
forest	257	3	0	42	0	0	65	1620	0	1987
driveway	0	0	0	0	0	92	1	0	158	251
	268	507	915	356	468	791	2060	1671	160	

Reference Source / # :	Independent - 7434	
Pixels Verified :	7196	
Diagonal Elements :	6637	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.992	(0.916 , 0.929)
Weighted Accuracy (w) :	0.848	(0.843 , 0.854)
Kappa ($\hat{k}$) :	0.905	(0.897 , 0.912)
B&P's Kappa ($\hat{k}_n$) :	0.913	(0.906 , 0.920)
Tau (T_p) :	0.906	(0.898 , 0.913)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	248095
Pixels Classified :	232241
Percent Classified :	0.936

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	250	0	0	6	0	0	0	0	0	256
sand	0	1504	0	0	0	0	0	0	0	1504
parking	0	0	199	0	0	0	0	0	0	199
dirt	0	0	0	40	0	0	0	0	0	40
roof	0	0	0	0	43	0	0	0	0	43
road	0	0	0	0	0	259	0	0	13	272
scrub	0	0	0	0	0	1	2043	0	0	2044
forest	0	0	0	0	0	0	23	2907	0	2930
driveway	0	0	0	0	0	1	0	0	72	73
	250	1504	199	46	43	261	2066	2907	85	

Reference Source / # :	Scaled Independent - 7434	
Pixels Verified :	7361	
Diagonal Elements :	7317	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.994	(0.992 , 0.996)
Weighted Accuracy (w) :	0.966	(0.953 , 0.980)
Kappa ($\hat{k}$) :	0.992	(0.989 , 0.994)
B&P's Kappa ($\hat{k}_n$) :	0.993	(0.991 , 0.995)
Tau (T_p) :	0.992	(0.989 , 0.994)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	248095
Pixels Classified :	228878
Percent Classified :	92.3

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	318	0	0	7	0	0	0	0	0	325
sand	0	1399	0	0	0	0	0	0	0	1399
parking	0	0	201	0	0	0	0	0	0	201
dirt	0	0	0	39	0	0	0	0	0	39
roof	0	0	0	0	44	0	0	0	0	44
road	0	0	0	0	0	270	0	0	11	281
scrub	0	0	0	0	0	1	2082	0	0	2083
forest	0	0	0	0	0	0	24	2921	0	2945
driveway	0	0	0	0	0	0	0	0	42	42
	318	1399	201	46	44	271	2106	2921	53	

Reference Source / # :	Scaled Independent - 7434	
Pixels Verified :	7359	
Diagonal Elements :	7316	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_0) :	0.994	(0.992 , 0.996)
Weighted Accuracy (w) :	0.958	(0.942 , 0.975)
Kappa ($\hat{k}$) :	0.992	(0.989 , 0.994)
B&P's Kappa ($\hat{k}_n$) :	0.993	(0.991 , 0.995)
Tau (T_p) :	0.992	(0.989 , 0.994)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	248095
Pixels Classified :	222617
Percent Classified :	89.7

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	254	0	0	4	0	0	0	0	0	258
sand	0	1468	0	0	0	0	0	0	0	1468
parking	0	0	200	0	0	0	0	0	0	200
dirt	0	0	0	43	0	0	0	0	0	43
roof	0	0	0	0	43	0	0	0	0	43
road	0	0	0	0	0	228	0	0	11	239
scrub	0	0	0	0	0	1	2104	0	0	2105
forest	0	0	0	0	0	0	23	2882	0	2905
driveway	0	0	0	0	0	1	0	0	75	76
	254	1468	200	47	43	230	2127	2882	86	

Reference Source / # :	Scaled Independent - 7434	
Pixels Verified :	7337	
Diagonal Elements :	7297	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.995	(0.993 , 0.996)
Weighted Accuracy (w) :	0.974	(0.962 , 0.986)
Kappa ($\hat{k}$) :	0.992	(0.990 , 0.995)
B&P's Kappa ($\hat{k}_n$) :	0.994	(0.992 , 0.996)
Tau (T_p) :	0.992	(0.990 , 0.995)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold 2 bands rule with large training set
Total Pixels :	248095
Pixels Classified :	262144
Percent Classified :	105.7

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	796	505	0	15	0	0	0	0	0	1316
sand	0	711	3	1	0	2	94	75	0	886
parking	0	2	258	0	1	1	0	0	0	262
dirt	41	0	0	60	0	0	0	0	0	101
roof	0	0	0	0	51	0	0	0	0	51
road	0	0	0	0	0	227	0	0	28	255
scrub	0	0	2	0	0	17	1387	0	5	1411
forest	0	0	0	0	0	1	79	3070	0	3150
driveway	0	0	0	0	0	0	0	0	2	2
	837	1218	263	76	52	248	1560	3145	35	

Reference Source / # :	Scaled Independent - 7434
Pixels Verified :	7434
Diagonal Elements :	6562

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.883	(0.875 , 0.890)
Weighted Accuracy (w) :	0.782	(0.776 , 0.807)
Kappa ($\hat{k}$) :	0.841	(0.832 , 0.851)
B&P's Kappa ($\hat{k}_n$) :	0.868	(0.860 , 0.876)
Tau (T_p) :	0.840	(0.830 , 0.850)

Scene Name : Tank Scene

Image Dimensions : 512 x 512 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 248095

Pixels Classified : 247396

Percent Classified : 99.7

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	0	0	0	0	0	0	0	0	0	0
sand	0	509	0	0	0	0	0	0	0	509
parking	0	0	356	0	0	8	0	0	2	366
dirt	0	1	0	598	0	0	0	0	0	599
roof	0	0	1	0	44	0	0	0	0	45
road	0	0	0	0	0	545	0	65	0	610
scrub	0	0	0	0	0	13	2233	10	0	2256
forest	0	3	0	80	0	0	73	2397	0	2553
driveway	0	0	0	0	0	74	1	0	180	255
	0	513	357	678	44	640	2307	2472	182	

Reference Source / # : Scaled Independent - 7434

Pixels Verified : 7196

Diagonal Elements : 6637

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.954	(0.949 , 0.959)
Weighted Accuracy (w) :	0.850	(0.845 , 0.855)
Kappa ($\hat{k}$) :	0.939	(0.933 , 0.945)
B&P's Kappa ($\hat{k}_n$) :	0.948	(0.943 , 0.954)
Tau (T_p) :	0.939	(0.933 , 0.945)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	248095
Pixels Classified :	232241
Percent Classified :	93.6

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	3	1	0	0	0	0	4	0	0	8
sand	0	39	0	0	0	0	10	1	0	50
parking	0	0	10	0	0	0	0	0	0	10
dirt	0	0	0	2	0	0	0	0	0	2
roof	0	0	0	0	1	0	0	0	0	1
road	0	2	2	0	0	5	0	0	0	9
scrub	0	1	0	0	0	0	60	7	0	68
forest	0	0	0	0	0	0	3	111	0	114
driveway	0	0	0	0	0	1	0	0	2	3
	3	43	12	2	1	6	77	119	2	

Reference Source / # :	Random Data Set - 278	
Pixels Verified :	265	
Diagonal Elements :	233	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_0) :	0.879	(0.840 , 0.918)
Weighted Accuracy (w) :	0.921	(0.877 , 0.964)
Kappa ($\hat{k}$) :	0.827	(0.772 , 0.882)
B&P's Kappa ($\hat{k}_n$) :	0.864	(0.820 , 0.908)
Tau (T_p) :	0.824	(0.766 , 0.881)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	248095
Pixels Classified :	228878
Percent Classified :	92.3

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	3	4	0	0	0	0	5	0	0	12
sand	0	36	0	0	0	0	9	1	0	46
parking	0	0	11	0	0	0	0	0	0	11
dirt	0	0	0	2	0	0	0	0	0	2
roof	0	0	0	0	1	0	0	0	0	1
road	0	2	1	0	0	5	0	0	0	8
scrub	0	1	0	0	0	0	60	6	0	67
forest	0	0	0	0	0	0	2	112	0	114
driveway	0	0	0	0	0	1	0	0	2	3
	3	43	12	2	1	6	76	119	2	

Reference Source / # :	Random Data Set - 278	
Pixels Verified :	264	
Diagonal Elements :	232	
	Coefficients	**95% Confidence Interval**
Simple Accuracy (P_o) :	0.879	(0.839 , 0.918)
Weighted Accuracy (w) :	0.924	(0.883 , 0.965)
Kappa ($\hat{k}$) :	0.827	(0.772 , 0.882)
B&P's Kappa ($\hat{k}_n$) :	0.864	(0.819 , 0.908)
Tau (T_p) :	0.823	(0.765 , 0.880)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	248095
Pixels Classified :	222617
Percent Classified :	89.7

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	3	0	0	0	0	0	4	0	0	7
sand	0	40	0	0	0	0	10	1	0	51
parking	0	0	11	0	0	0	0	0	0	11
dirt	0	0	0	2	0	0	0	0	0	2
roof	0	0	0	0	1	0	0	0	0	1
road	0	1	1	0	0	5	0	0	0	7
scrub	0	1	0	0	0	0	60	7	0	68
forest	0	0	0	0	0	0	1	105	0	106
driveway	0	0	0	0	0	1	0	0	2	3
	3	42	12	2	1	6	75	113	2	

	Coefficients	95% Confidence Interval
Reference Source / # :	Random - 278	
Pixels Verified :	256	
Diagonal Elements :	229	
Simple Accuracy (P_o) :	0.895	(0.857 , 0.932)
Weighted Accuracy (w) :	0.937	(0.897 , 0.977)
Kappa ($\hat{k}$) :	0.850	(0.797 , 0.903)
B&P's Kappa ($\hat{k}_n$) :	0.881	(0.839 , 0.924)
Tau (T_p) :	0.847	(0.792 , 0.902)

Scene Name :	Tank Scene
Image Dimensions :	512 x 512 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold 2 bands rule with large training set
Total Pixels :	248095
Pixels Classified :	262144
Percent Classified :	105.7

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	3	23	0	0	0	0	1	0	0	27
sand	0	18	1	0	0	0	21	5	0	45
parking	0	0	11	0	0	0	1	0	0	12
dirt	0	1	0	2	0	0	0	0	0	3
roof	0	0	0	0	2	0	0	0	0	2
road	0	0	0	0	0	5	1	1	1	8
scrub	0	3	1	0	0	1	51	5	0	61
forest	0	0	0	0	0	0	5	112	0	117
driveway	0	0	0	0	0	0	2	0	1	3
	3	45	13	2	2	6	82	123	2	

Reference Source / # :	Random - 278
Pixels Verified :	278
Diagonal Elements :	205

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.737	(0.686 , 0.789)
Weighted Accuracy (w) :	0.790	(0.701 , 0.879)
Kappa ($\hat{k}$) :	0.635	(0.568 , 0.701)
B&P's Kappa ($\hat{k}_n$) :	0.705	(0.646 , 0.763)
Tau (T_p) :	0.618	(0.543 , 0.694)

Scene Name : Tank Scene

Image Dimensions : 512 x 512 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 248095

Pixels Classified : 247396

Percent Classified : 99.7

Confusion Matrix

	trail	sand	parking	dirt	roof	road	scrub	forest	driveway	
trail	0	0	0	0	0	0	0	0	0	0
sand	0	18	1	0	0	0	0	1	0	20
parking	0	2	11	0	0	0	1	0	0	14
dirt	0	16	0	2	0	0	7	0	0	25
roof	0	0	0	0	1	0	0	0	0	1
road	0	0	0	0	0	4	2	23	0	29
scrub	0	4	1	0	0	0	68	13	0	86
forest	3	5	0	0	0	0	1	85	0	94
driveway	0	0	0	0	0	2	3	1	2	8
	3	45	13	2	1	6	82	123	2	

Reference Source / # : Random - 278

Pixels Verified : 277

Diagonal Elements : 191

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.690	(0.635 , 0.744)
Weighted Accuracy (w) :	0.715	(0.663 , 0.766)
Kappa ($\hat{k}$) :	0.581	(0.513 , 0.648)
B&P's Kappa ($\hat{k}_n$) :	0.651	(0.589 , 0.712)
Tau (T_p) :	0.547	(0.468 , 0.627)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.0 Probability Threshold
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	576	0	0	0	0	576
road	0	204	0	0	0	204
grass	0	17	277	0	3	297
metal	0	0	0	28	0	28
water	0	0	0	0	99	99
	576	221	277	28	102	

Reference Source / # :	Dependent - 1204	
Pixels Verified :	1204	
Diagonal Elements :	1184	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.983	(0.976 , 0.991)
Weighted Accuracy (w) :	0.979	(0.957 , 1.000)
Kappa ($\hat{k}$) :	0.975	(0.965 , 0.986)
B&P's Kappa ($\hat{k}_n$) :	0.979	(0.970 , 0.988)
Tau (T_p) :	0.975	(0.965 , 0.986)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	408	171	38	0	0	617
road	0	0	0	0	0	0
grass	13	27	239	20	3	302
metal	0	0	0	5	1	6
water	155	23	0	3	98	279
	576	221	277	28	102	

Reference Source / # :	Dependent - 1204
Pixels Verified :	1204
Diagonal Elements :	750

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.623	(0.596 , 0.650)
Weighted Accuracy (w) :	0.542	(0.472 , 0.612)
Kappa ($\hat{k}$) :	0.443	(0.406 , 0.480)
B&P's Kappa ($\hat{k}_n$) :	0.529	(0.494 , 0.563)
Tau (T_p) :	0.443	(0.402 , 0.483)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	576	0	0	0	0	576
road	0	213	42	0	0	255
grass	0	8	235	27	0	270
metal	0	0	0	0	0	0
water	0	0	0	1	102	103
	576	221	277	28	102	

Reference Source / # :	Dependent - 1204	
Pixels Verified :	1204	
Diagonal Elements :	1126	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.935	(0.921 , 0.949)
Weighted Accuracy (w) :	0.762	(0.741 , 0.784)
Kappa ($\hat{k}$) :	0.904	(0.884 , 0.924)
B&P's Kappa ($\hat{k}_n$) :	0.919	(0.902 , 0.936)
Tau (T_p) :	0.904	(0.884 , 0.925)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 7km Visibility

Classifier : Gaussian Maximum Likelihood

Notes : 0.0 Probability Threshold

Total Pixels : 62500

Pixels Classified :62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	576	0	0	0	0	576
road	0	179	0	0	0	179
grass	0	42	277	0	2	321
metal	0	0	0	28	0	28
water	0	0	0	0	100	100
	576	221	277	28	102	

Reference Source / # : Dependent - 1204

Pixels Verified : 1204

Diagonal Elements : 1160

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.963	(0.953 , 0.974)
Weighted Accuracy (w) :	0.958	(0.932 , 0.984)
Kappa ($\hat{k}$) :	0.946	(0.930 , 0.961)
B&P's Kappa ($\hat{k}_n$) :	0.954	(0.941 , 0.968)
Tau (T_p) :	0.946	(0.930 , 0.962)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	570	150	127	2	1	850
road	0	69	0	0	0	69
grass	4	2	150	17	3	176
metal	2	0	0	9	1	12
water	0	0	0	0	97	97
	576	221	277	28	102	

Reference Source / # :	Dependent - 1204
Pixels Verified :	1204
Diagonal Elements :	895

	Coefficients	95% Confidence Interval
Simple Accuracy (P_0) :	0.743	(0.719 , 0.768)
Weighted Accuracy (w) :	0.623	(0.535 , 0.711)
Kappa ($\hat{k}$) :	0.580	(0.542 , 0.618)
B&P's Kappa ($\hat{k}_n$) :	0.679	(0.648 , 0.710)
Tau (T_p) :	0.621	(0.584 , 0.657)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 7km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	576	0	0	0	0	576
road	0	221	104	0	0	325
grass	0	0	173	28	0	201
metal	0	0	0	0	0	0
water	0	0	0	0	102	102
	576	221	277	28	102	

Reference Source / # : Dependent - 1204

Pixels Verified : 1204

Diagonal Elements : 1072

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.890	(0.873 , 0.908)
Weighted Accuracy (w) :	0.725	(0.699 , 0.750)
Kappa ($\hat{k}$) :	0.838	(0.813 , 0.863)
B&P's Kappa ($\hat{k}_n$) :	0.863	(0.841 , 0.885)
Tau (T_p) :	0.838	(0.812 , 0.864)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Gaussian Maximum Likelihood

Notes : 0.0 Probability Threshold

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	576	0	0	0	0	576
road	0	200	0	0	1	201
grass	0	21	235	0	2	258
metal	0	0	0	25	0	25
water	0	0	42	3	99	144
	576	221	277	28	102	

Reference Source / # : Dependent - 1204

Pixels Verified : 1204

Diagonal Elements : 1135

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.943	(0.930 , 0.956)
Weighted Accuracy (w) :	0.923	(0.864 , 0.982)
Kappa ($\hat{k}$) :	0.916	(0.897 , 0.935)
B&P's Kappa ($\hat{k}_n$) :	0.928	(0.912 , 0.945)
Tau (T_p) :	0.915	(0.896 , 0.935)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Rule Based Genetic Algorithm (MYSTIC)

Notes : threshold two bands, large training set

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	528	2	0	1	0	531
road	0	175	0	0	0	175
grass	45	44	277	15	4	385
metal	3	0	0	12	0	15
water	0	0	0	0	98	98
	576	221	277	28	102	

Reference Source / # : Dependent - 1204

Pixels Verified : 1204

Diagonal Elements : 1090

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.905	(0.889 , 0.922)
Weighted Accuracy (w) :	0.820	(0.732 , 0.907)
Kappa ($\hat{k}$) :	0.861	(0.837 , 0.885)
B&P's Kappa ($\hat{k}_n$) :	0.882	(0.861 , 0.902)
Tau (T_p) :	0.860	(0.836 , 0.885)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	574	10	0	0	0	584
road	2	211	63	0	0	276
grass	0	0	214	28	0	242
metal	0	0	0	0	0	0
water	0	0	0	0	102	102
	576	221	277	28	102	

Reference Source / # :	Dependent - 1204	
Pixels Verified :	1204	
Diagonal Elements :	1101	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.914	(0.899 , 0.930)
Weighted Accuracy (w) :	0.745	(0.719 , 0.770)
Kappa ($\hat{k}$) :	0.873	(0.850 , 0.896)
B&P's Kappa ($\hat{k}_n$) :	0.893	(0.873 , 0.913)
Tau (T_p) :	0.874	(0.850 , 0.897)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.0 Probability Threshold
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	269	0	0	20	0	289
road	0	179	11	6	0	196
grass	1	4	1866	1	0	1872
metal	0	0	0	53	0	53
water	0	0	0	0	302	302
	270	183	1877	80	302	

Reference Source / # :	Independent - 2712
Pixels Verified :	2712
Diagonal Elements :	2669

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.984	(0.979 , 0.989)
Weighted Accuracy (w) :	0.926	(0.879 , 0.974)
Kappa ($\hat{k}$) :	0.968	(0.958 , 0.977)
B&P's Kappa ($\hat{k}_n$) :	0.980	(0.974 , 0.986)
Tau (T_p) :	0.968	(0.958 , 0.977)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	206	182	509	1	0	898
road	0	0	0	0	0	0
grass	4	1	1367	26	0	1398
metal	0	0	0	7	1	8
water	60	0	1	46	301	408
	270	183	1877	80	302	

Reference Source / # :	Independent - 2712
Pixels Verified :	2712
Diagonal Elements :	1881

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.694	(0.676 , 0.711)
Weighted Accuracy (w) :	0.515	(0.478 , 0.552)
Kappa ($\hat{k}$) :	0.484	(0.459 , 0.508)
B&P's Kappa ($\hat{k}_n$) :	0.617	(0.595 , 0.639)
Tau (T_p) :	0.379	(0.344 , 0.414)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	267	177	5	43	0	492
road	3	5	707	4	0	719
grass	0	1	1100	8	0	1109
metal	0	0	0	0	0	0
water	0	0	65	25	302	392
	270	183	1877	80	302	

Reference Source / # :	Independent - 2712
Pixels Verified :	2712
Diagonal Elements :	1674

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.617	(0.599 , 0.636)
Weighted Accuracy (w) :	0.520	(0.505 , 0.536)
Kappa ($\hat{k}$) :	0.424	(0.401 , 0.448)
B&P's Kappa ($\hat{k}_n$) :	0.552	(0.499 , 0.544)
Tau (T_p) :	0.224	(0.187 , 0.261)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.0 Probability Threshold
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	270	0	0	3	0	273
road	0	162	27	0	0	189
grass	0	21	1850	1	0	1872
metal	0	0	0	76	9	85
water	0	0	0	0	293	293
	270	183	1877	80	302	

Reference Source / # :	Independent - 2712	
Pixels Verified :	2712	
Diagonal Elements :	2651	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_0) :	0.978	(0.972 , 0.983)
Weighted Accuracy (w) :	0.958	(0.927 , 0.989)
Kappa ($\hat{k}$) :	0.955	(0.943 , 0.966)
B&P's Kappa ($\hat{k}_n$) :	0.972	(0.965 , 0.979)
Tau (T_p) :	0.954	(0.943 , 0.966)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 7km Visibility

Classifier : Rule Based Genetic Algorithm (MYSTIC)

Notes : threshold two bands, large training set

Total Pixels : 62500

Pixels Classified :62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	268	126	836	5	1	1236
road	0	57	0	0	0	57
grass	1	0	1041	0	0	1042
metal	1	0	0	75	0	76
water	0	0	0	0	301	301
	270	183	1877	80	302	

Reference Source / # : Independent - 2712

Pixels Verified : 2712

Diagonal Elements : 1742

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.642	(0.624 , 0.660)
Weighted Accuracy (w) :	0.759	(0.719 , 0.798)
Kappa ($\hat{k}$) :	0.469	(0.445 , 0.493)
B&P's Kappa ($\hat{k}_n$) :	0.553	(0.530 , 0.575)
Tau (T_p) :	0.275	(0.238 , 0.311)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	270	0	4	7	0	281
road	0	176	869	0	0	1045
grass	0	7	823	37	0	867
metal	0	0	0	0	0	0
water	0	0	181	36	302	519
	270	183	1877	80	302	

Reference Source / # :	Independent - 2712	
Pixels Verified :	2712	
Diagonal Elements :	1571	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.579	(0.561 , 0.598)
Weighted Accuracy (w) :	0.680	(0.664 , 0.696)
Kappa ($\hat{k}$) :	0.417	(0..393 , 0.440)
B&P's Kappa ($\hat{k}_n$) :	0.474	(0.451 , 0.497)
Tau (T_p) :	0.147	(0.109 , 0.185)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.0 Probability Threshold
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	270	0	0	0	46	316
road	0	177	9	0	0	186
grass	0	6	1699	1	0	1706
metal	0	0	0	75	0	75
water	0	0	169	4	256	429
	270	183	1877	80	302	

Reference Source / # :	Independent - 2712
Pixels Verified :	2712
Diagonal Elements :	2477

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.913	(0.903 , 0.924)
Weighted Accuracy (w) :	0.932	(0.899 , 0.964)
Kappa ($\hat{k}$) :	0.836	(0.817 , 0.856)
B&P's Kappa ($\hat{k}_n$) :	0.892	(0.878 , 0.905)
Tau (T_p) :	0.824	(0.803 , 0.846)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	ratio two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	264	0	0	0	0	264
road	0	164	6	0	0	170
grass	4	19	1871	56	16	1966
metal	2	0	0	24	0	26
water	0	0	0	0	286	286
	270	183	1877	80	302	

Reference Source / # :	Independent - 2712
Pixels Verified :	2712
Diagonal Elements :	2609

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.962	(0.955 , 0.969)
Weighted Accuracy (w) :	0.824	(0.773 , 0.875)
Kappa ($\hat{k}$) :	0.920	(0.904 , 0.935)
B&P's Kappa ($\hat{k}_n$) :	0.953	(0.944 , 0.962)
Tau (T_p) :	0.923	(0.908 , 0.938)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	268	9	33	0	0	310
road	1	174	367	1	0	543
grass	1	0	1477	46	0	1524
metal	0	0	0	0	0	0
water	0	0	0	33	302	335
	270	183	1877	80	302	

Reference Source / # :	Independent - 2712	
Pixels Verified :	2712	
Diagonal Elements :	2221	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.819	(0.804 , 0.833)
Weighted Accuracy (w) :	0.746	(0.729 , 0.763)
Kappa ($\hat{k}$) :	0.684	(0.659 , 0.708)
B&P's Kappa ($\hat{k}_n$) :	0.774	(0.756 , 0.792)
Tau (T_p) :	0.633	(0.604 , 0.662)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.0 Probability Threshold
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

$$
\begin{array}{c}
\\
\end{array}
\left(
\begin{array}{lcccccc}
 & \text{forest} & \text{road} & \text{grass} & \text{metal} & \text{water} & \\
\text{forest} & 1506 & 0 & 0 & 16 & 0 & 1522 \\
\text{road} & 0 & 382 & 4 & 5 & 0 & 391 \\
\text{grass} & 6 & 9 & 702 & 1 & 0 & 718 \\
\text{metal} & 0 & 0 & 0 & 41 & 0 & 41 \\
\text{water} & 0 & 0 & 0 & 0 & 41 & 41 \\
 & & & & & & \\
 & 1512 & 391 & 706 & 63 & 41 &
\end{array}
\right)
$$

Reference Source / # :	Scaled Independent - 2712
Pixels Verified :	2713
Diagonal Elements :	2672

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.985	(0.980 , 0.989)
Weighted Accuracy (w) :	0.924	(0.870 , 0.977)
Kappa ($\hat{k}$) :	0.975	(0.967 , 0.982)
B&P's Kappa ($\hat{k}_n$) :	0.981	(0.975 , 0.987)
Tau (T_p) :	0.975	(0.967 , 0.982)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	864	0	214	0	0	1078
road	0	0	0	0	0	0
grass	17	0	576	11	0	604
metal	0	0	0	3	2	5
water	252	0	0	20	751	1023
	1133	0	790	34	753	

Reference Source / # :	Scaled Independent - 2712	
Pixels Verified :	2710	
Diagonal Elements :	2194	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.810	(0.795 , 0.824)
Weighted Accuracy (w) :	0.515	(0.469 , 0.562)
Kappa ($\hat{k}$) :	0.713	(0.691 , 0.735)
B&P's Kappa ($\hat{k}_n$) :	0.762	(0.744 , 0.780)
Tau (T_p) :	0.713	(0.690 , 0.735)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	1268	831	1	0	0	2100
road	14	23	160	0	0	197
grass	0	5	250	0	0	255
metal	0	0	0	0	0	0
water	0	0	15	0	144	159
	1282	859	426	0	144	

Reference Source / # :	Scaled Independent - 2712
Pixels Verified :	2711
Diagonal Elements :	1685

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.622	(0.603 , 0.640)
Weighted Accuracy (w) :	0.521	(0.499 , 0.542)
Kappa ($\hat{k}$) :	0.362	(0.335 , 0. 388)
B&P's Kappa ($\hat{k}_n$) :	0.527	(0.504 , 0.550)
Tau (T_p) :	0.416	(0.388 , 0.445)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.0 Probability Threshold
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	1701	0	0	1	0	1702
road	0	130	12	0	0	142
grass	0	17	795	0	0	812
metal	0	0	0	14	1	15
water	0	0	0	0	41	41
	1701	147	807	15	42	

Reference Source / # :	Scaled Independent - 2712
Pixels Verified :	2712
Diagonal Elements :	2681

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.989	(0.985 , 0.993)
Weighted Accuracy (w) :	0.956	(0.921 , 0.990)
Kappa ($\hat{k}$) :	0.978	(0.970 , 0.985)
B&P's Kappa ($\hat{k}_n$) :	0.986	(0.981 , 0.991)
Tau (T_p) :	0.978	(0.970 , 0.986)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 7km Visibility

Classifier : Rule Based Genetic Algorithm (MYSTIC)

Notes : threshold two bands, large training set

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	2016	14	266	1	0	2297
road	0	6	0	0	0	6
grass	8	0	331	0	0	339
metal	8	0	0	22	0	30
water	0	0	0	0	41	41
	2032	20	597	23	41	

Reference Source / # : Scaled Independent - 2712

Pixels Verified : 2713

Diagonal Elements : 2416

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.891	(0.879 , 0.902)
Weighted Accuracy (w) :	0.761	(0.662 , 0.859)
Kappa ($\hat{k}$) :	0.676	(0.643 , 0.710)
B&P's Kappa ($\hat{k}_n$) :	0.863	(0.848 , 0.878)
Tau (T_p) :	0.719	(0.689 , 0.750)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

$$\left(\begin{array}{lrrrrrr} & \text{forest} & \text{road} & \text{grass} & \text{metal} & \text{water} & \\ \text{forest} & 1750 & 0 & 1 & 0 & 0 & 1751 \\ \text{road} & 0 & 353 & 201 & 0 & 0 & 554 \\ \text{grass} & 0 & 14 & 190 & 0 & 0 & 204 \\ \text{metal} & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{water} & 0 & 0 & 42 & 0 & 162 & 204 \\ & & & & & & \\ & 1750 & 367 & 434 & 0 & 162 & \end{array}\right)$$

Reference Source / # :	Scaled Independent - 2712	
Pixels Verified :	2713	
Diagonal Elements :	2455	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.905	(0.894 , 0.916)
Weighted Accuracy (w) :	0.680	(0.657 , 0.703)
Kappa ($\hat{k}$) :	0.824	(0.805 , 0.842)
B&P's Kappa ($\hat{k}_n$) :	0.881	(0.867 , 0.895)
Tau (T_p) :	0.823	(0.802 , 0.843)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.0 Probability Threshold
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	1685	0	0	0	16	1701
road	0	196	3	0	0	199
grass	0	7	626	0	0	633
metal	0	0	0	27	0	27
water	0	0	62	1	88	151
	1685	203	691	28	104	

Reference Source / # :	Scaled Independent - 2712
Pixels Verified :	2711
Diagonal Elements :	2622

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.967	(0.960 , 0.974)
Weighted Accuracy (w) :	0.936	(0.890 , 0.983)
Kappa ($\hat{k}$) :	0.940	(0.927 , 0.952)
B&P's Kappa ($\hat{k}_n$) :	0.959	(0.951 , 0.967)
Tau (T_p) :	0.939	(0.927 , 0.952)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	ratio two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	1372	0	0	0	0	1372
road	0	84	3	0	0	87
grass	21	10	1042	90	2	1165
metal	10	0	0	39	0	49
water	0	0	0	0	38	38
	1403	94	1045	129	40	

Reference Source / # :	Scaled Independent - 2712	
Pixels Verified :	2711	
Diagonal Elements :	2575	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.950	(0.942 , 0.958)
Weighted Accuracy (w) :	0.824	(0.770 , 0.879)
Kappa ($\hat{k}$) :	0.912	(0.898 , 0.926)
B&P's Kappa ($\hat{k}_n$) :	0.937	(0.927, 0.948)
Tau (T_p) :	0.913	(0.899 , 0.928)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	1507	17	13	0	0	1537
road	6	338	145	0	0	489
grass	6	0	583	0	0	589
metal	0	0	0	0	0	0
water	0	0	0	0	97	97
	1519	355	741	0	97	

Reference Source / # : Scaled Independent - 2712

Pixels Verified : 2712

Diagonal Elements : 2525

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.931	(0.922 , 0.941)
Weighted Accuracy (w) :	0.746	(0.730 , 0.763)
Kappa ($\hat{k}$) :	0.885	(0.869, 0.900)
B&P's Kappa ($\hat{k}_n$) :	0.914	(0.902 , 0.926)
Tau (T_p) :	0.884	(0.868 , 0.900)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 5km Visibility

Classifier : Gaussian Maximum Likelihood

Notes : 0.0 Probability Threshold

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	145	2	19	1	0	167
road	22	8	10	0	0	40
grass	1	0	78	0	0	79
metal	3	0	3	1	0	7
water	0	0	0	0	7	7
	171	10	110	2	7	

Reference Source / # : Random - 300

Pixels Verified : 300

Diagonal Elements : 239

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.797	(0.751 , 0.842)
Weighted Accuracy (w) :	0.771	(0.623 , 0.920)
Kappa ($\hat{k}$) :	0.650	(0.577 , 0.723)
B&P's Kappa ($\hat{k}_n$) :	0.746	(0.689 , 0.803)
Tau (T_p) :	0.623	(0.538 , 0.707)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	94	6	27	0	0	127
road	0	0	0	0	0	0
grass	17	0	68	0	0	85
metal	4	0	0	0	0	4
water	56	4	15	2	7	84
	171	10	110	2	7	

Reference Source / # :	Random - 300
Pixels Verified :	300
Diagonal Elements :	169

	Coefficients	95% Confidence Interval
Simple Accuracy (P_0) :	0.563	(0.507 , 0.619)
Weighted Accuracy (w) :	0.434	(0.410 , 0.457)
Kappa ($\hat{k}$) :	0.326	(0.252 , 0.401)
B&P's Kappa ($\hat{k}_n$) :	0.454	(0.384 , 0.524)
Tau (T_p) :	0.190	(0.086 , 0.294)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	108	5	28	2	0	143
road	46	5	43	0	0	94
grass	15	0	32	0	0	47
metal	0	0	0	0	0	0
water	2	0	7	0	7	16
	171	10	110	2	7	

Reference Source / # :	Random - 300	
Pixels Verified :	300	
Diagonal Elements :	152	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.507	(0.450 , 0.563)
Weighted Accuracy (w) :	0.484	(0.419 , 0.550
Kappa ($\hat{k}$) :	0.252	(0.181 , 0.322)
B&P's Kappa ($\hat{k}_n$) :	0.383	(0.313 , 0.454)
Tau (T_p) :	0.085	(0.000 , 0.190)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.0 Probability Threshold
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	165	1	25	1	0	192
road	5	8	1	0	0	14
grass	1	1	84	0	0	86
metal	0	0	0	1	0	1
water	0	0	0	0	7	7
	171	10	110	2	7	

Reference Source / # :	Random - 300	
Pixels Verified :	300	
Diagonal Elements :	265	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.883	(0.847 , 0.920)
Weighted Accuracy (w) :	0.806	(0.658 , 0.954)
Kappa ($\hat{k}$) :	0.779	(0.711 , 0.847)
B&P's Kappa ($\hat{k}_n$) :	0.854	(0.809 , 0.900)
Tau (T_p) :	0.784	(0.716 , 0.851)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	169	8	49	1	0	227
road	0	2	0	0	0	2
grass	1	0	61	0	0	62
metal	1	0	0	1	0	2
water	0	0	0	0	7	7
	171	10	110	2	7	

Reference Source / # :	Random - 300	
Pixels Verified :	300	
Diagonal Elements :	240	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.800	(0.755 , 0.845)
Weighted Accuracy (w) :	0.649	(0.500 , 0.797)
Kappa ($\hat{k}$) :	0.594	(0.507 , 0.680)
B&P's Kappa ($\hat{k}_n$) :	0.750	(0.693 , 0.807)
Tau (T_p) :	0.629	(0.545 , 0.713)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	168	2	24	1	0	195
road	2	8	34	0	0	44
grass	0	0	36	0	0	36
metal	0	0	0	0	0	0
water	1	0	16	1	7	25
	171	10	110	2	7	

Reference Source / # :	Random - 300	
Pixels Verified :	300	
Diagonal Elements :	219	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.730	(0.680 , 0.780)
Weighted Accuracy (w) :	0.622	(0.569 , 0.675)
Kappa ($\hat{k}$) :	0.533	(0.467 , 0.600)
B&P's Kappa ($\hat{k}_n$) :	0.663	(0.600 , 0.725)
Tau (T_p) :	0.499	(0.406 , 0.592)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.0 Probability Threshold
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	163	0	24	0	0	187
road	7	9	4	0	0	20
grass	1	0	78	0	0	79
metal	0	1	0	2	0	3
water	0	0	4	0	7	11
	171	10	110	2	7	

Reference Source / # :	Random - 300
Pixels Verified :	300
Diagonal Elements :	259

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.863	(0.824 , 0.902)
Weighted Accuracy (w) :	0.912	(0.871 , 0.954)
Kappa ($\hat{k}$) :	0.749	(0.680 , 0.818)
B&P's Kappa ($\hat{k}_n$) :	0.829	(0.781 , 0.878)
Tau (T_p) :	0.746	(0.674 , 0.819)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Rule Based Genetic Algorithm (MYSTIC)

Notes : ratio two bands, large training set

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	159	0	8	0	0	167
road	0	7	1	0	0	8
grass	8	2	98	1	0	109
metal	4	1	3	1	0	9
water	0	0	0	0	7	7
	171	10	110	2	7	

Reference Source / # : Random - 300

Pixels Verified : 300

Diagonal Elements : 272

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.907	(0.874 , 0.940)
Weighted Accuracy (w) :	0.804	(0.654 , 0.955)
Kappa ($\hat{k}$) :	0.830	(0.771 , 0.888)
B&P's Kappa ($\hat{k}_n$) :	0.883	(0.842 , 0.924)
Tau (T_p) :	0.827	(0.766 , 0.888)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	149	2	21	0	0	172
road	8	7	23	0	0	38
grass	8	1	66	2	0	77
metal	0	0	0	0	0	0
water	6	0	0	0	7	13
	171	10	110	2	7	

Reference Source / # :	Random - 300	
Pixels Verified :	300	
Diagonal Elements :	229	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.763	(0.715 , 0.811)
Weighted Accuracy (w) :	0.634	(0.574 , 0.695)
Kappa ($\hat{k}$) :	0.588	(0.511 , 0.664)
B&P's Kappa ($\hat{k}_n$) :	0.704	(0.644 , 0.764)
Tau (T_p) :	0.561	(0.472 , 0.650)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 5km Visibility

Classifier : Gaussian Maximum Likelihood

Notes : 0.0 Probability Threshold

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	29211	416	5119	72	16	34834
road	2793	2283	3903	7	17	9003
grass	110	494	15650	3	12	16269
metal	541	14	750	129	4	1438
water	0	0	9	0	947	956
	32655	3207	25431	211	996	

Reference Source / # : Synthetic - 62500

Pixels Verified : 62500

Diagonal Elements : 48220

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.772	(0.772 , 0.772)
Weighted Accuracy (w) :	0.757	(0.757 , 0.757)
Kappa ($\hat{k}$) :	0.616	(0.616 , 0.616)
B&P's Kappa ($\hat{k}_n$) :	0.714	(0.714 , 0.714)
Tau (T_p) :	0.591	(0.591 , 0.591)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 5km Visibility

Classifier : Rule Based Genetic Algorithm (MYSTIC)

Notes : threshold two bands, large training set

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	18859	2095	5144	5	0	26103
road	0	0	0	0	0	0
grass	2303	259	15580	64	23	18229
metal	656	8	104	25	7	800
water	10837	845	4603	117	966	17368
	32655	3207	25431	211	996	

Reference Source / # : Synthetic - 62500

Pixels Verified : 62500

Diagonal Elements : 35430

	Coefficients	95% Confidence Interval
Simple Accuracy (P_0) :	0.567	(0.567 , 0.567)
Weighted Accuracy (w) :	0.456	(0.456 , 0.456)
Kappa ($\hat{k}$) :	0.342	(0.342 , 0.342)
B&P's Kappa ($\hat{k}_n$) :	0.459	(0.459 , 0.459)
Tau (T_p) :	0.225	(0.225 , 0.225)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	21137	1753	6535	124	0	29549
road	9291	1315	9186	16	1	19809
grass	1338	125	8310	41	3	9817
metal	0	0	0	0	0	0
water	889	14	1400	30	992	3325
	32655	3207	25431	211	996	

Reference Source / # :	Synthetic - 62500	
Pixels Verified :	62500	
Diagonal Elements :	31754	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.508	(0.508 , 0.508)
Weighted Accuracy (w) :	0.476	(0.476 , 0.476)
Kappa ($\hat{k}$) :	0.268	(0.268 , 0.268)
B&P's Kappa ($\hat{k}_n$) :	0.385	(0.385 , 0.385)
Tau (T_p) :	0.119	(0.119 , 0.119)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 7km Visibility

Classifier : Gaussian Maximum Likelihood

Notes : 0.0 Probability Threshold

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	32476	208	6467	44	9	39204
road	124	2440	828	4	1	3397
grass	35	546	18011	0	2	18594
metal	20	13	110	163	34	340
water	0	0	15	0	950	965
	32655	3207	25431	211	996	

Reference Source / # : Synthetic - 62500

Pixels Verified : 62500

Diagonal Elements : 54040

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.865	(0.865 , 0.865)
Weighted Accuracy (w) :	0.838	(0.838 , 0.838)
Kappa ($\hat{k}$) :	0.753	(0.753 , 0.753)
B&P's Kappa ($\hat{k}_n$) :	0.831	(0.831 , 0.831)
Tau (T_p) :	0.758	(0.758 , 0.758)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	32163	2769	11842	32	8	46814
road	52	405	6	0	0	463
grass	167	11	13533	25	9	13745
metal	273	22	50	154	34	533
water	0	0	0	0	945	945
	32655	3207	25431	211	996	

Reference Source / # :	Synthetic - 62500	
Pixels Verified :	62500	
Diagonal Elements :	47200	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.755	(0.755 , 0.755)
Weighted Accuracy (w) :	0.664	(0.664 , 0.664)
Kappa ($\hat{k}$) :	0.528	(0.528 , 0.528)
B&P's Kappa ($\hat{k}_n$) :	0.694	(0.694 , 0.694)
Tau (T_p) :	0.562	(0.562 , 0.562)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	32311	1015	6950	41	2	40319
road	197	2063	6205	0	1	8466
grass	96	121	9668	78	24	9987
metal	0	0	0	0	0	0
water	51	8	2608	92	969	3728
	32655	3207	25431	211	996	

Reference Source / # :	Synthetic - 62500
Pixels Verified :	62500
Diagonal Elements :	45011

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.720	(0.720 , 0.720)
Weighted Accuracy (w) :	0.597	(0.597 , 0.597)
Kappa ($\hat{k}$) :	0.526	(0.526 , 0.526)
B&P's Kappa ($\hat{k}_n$) :	0.650	(0.650 , 0.650)
Tau (T_p) :	0.499	(0.499 , 0.499)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.0 Probability Threshold
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	31603	402	6693	46	86	38830
road	901	2327	1447	3	2	4680
grass	25	456	15446	7	4	15938
metal	120	18	354	152	8	652
water	6	4	1491	3	896	2400
	32655	3207	25431	211	996	

Reference Source / # :	Synthetic - 62500
Pixels Verified :	62500
Diagonal Elements :	50424

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.807	(0.807 , 0.807)
Weighted Accuracy (w) :	0.784	(0.784 , 0.784)
Kappa ($\hat{k}$) :	0.659	(0.659 , 0.659)
B&P's Kappa ($\hat{k}_n$) :	0.758	(0.758 , 0.758)
Tau (T_p) :	0.654	(0.654 , 0.654)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	ratio two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	31047	97	1194	5	0	32343
road	10	1945	195	0	0	2150
grass	903	926	22076	130	61	24096
metal	695	239	1962	76	7	2979
water	0	0	4	0	928	932
	32655	3207	25431	211	996	

Reference Source / # :	Synthetic - 62500
Pixels Verified :	62500
Diagonal Elements :	56072

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.897	(0.897 , 0.897)
Weighted Accuracy (w) :	0.743	(0.743 , 0.743)
Kappa ($\hat{k}$) :	0.820	(0.820 , 0.820)
B&P's Kappa ($\hat{k}_n$) :	0.871	(0.871 , 0.871)
Tau (T_p) :	0.816	(0.816 , 0.816)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	28569	1189	5217	18	0	34993
road	1188	1746	5260	0	0	8194
grass	1980	271	14658	148	16	17073
metal	0	0	0	0	0	0
water	918	1	296	45	980	2240
	32655	3207	25431	211	996	

Reference Source / # : Synthetic - 62500

Pixels Verified : 62500

Diagonal Elements : 45953

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.735	(0.735 , 0.735)
Weighted Accuracy (w) :	0.596	(0.596 , 0.596)
Kappa ($\hat{k}$) :	0.551	(0.551 , 0.551)
B&P's Kappa ($\hat{k}_n$) :	0.669	(0.669 , 0.669)
Tau (T_p) :	0.526	(0.526 , 0.526)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 5km Visibility

Classifier : Gaussian Maximum Likelihood

Notes : 0.0 Probability Threshold

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	523	0	53	0	0	576
road	0	170	34	0	0	204
grass	0	4	293	0	0	297
metal	0	2	0	26	0	28
water	0	0	1	0	98	99
	523	176	381	26	98	

Reference Source / # : Dependent/Synthetic - 1204

Pixels Verified : 1204

Diagonal Elements : 1110

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.922	(0.907 , 0.937)
Weighted Accuracy (w) :	0.947	(0.925 , 0.969)
Kappa ($\hat{k}$) :	0.886	(0.863 , 0.908)
B&P's Kappa ($\hat{k}_n$) :	0.902	(0.883 , 0.921)
Tau (T_p) :	0.886	(0.863 , 0.908)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	390	150	77	0	0	617
road	0	0	0	0	0	0
grass	8	4	271	19	0	302
metal	0	0	1	5	0	6
water	125	22	32	2	98	279
	523	176	381	26	98	

Reference Source / # :	Dependent/Synthetic - 1204	
Pixels Verified :	1204	
Diagonal Elements :	764	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.635	(0.607 , 0.662)
Weighted Accuracy (w) :	0.530	(0.457 , 0.603)
Kappa ($\hat{k}$) :	0.462	(0.425 , 0.498)
B&P's Kappa ($\hat{k}_n$) :	0.543	(0.509 , 0.577)
Tau (T_p) :	0.465	(0.425 , 0.505)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	523	0	53	0	0	576
road	0	170	85	0	0	255
grass	0	6	239	25	0	270
metal	0	0	0	0	0	0
water	0	0	4	1	98	103
	523	176	381	26	98	

Reference Source / # :	Dependent/Synthetic - 1204
Pixels Verified :	1204
Diagonal Elements :	1030

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.855	(0.836 , 0.875)
Weighted Accuracy (w) :	0.719	(0.694 , 0.743)
Kappa ($\hat{k}$) :	0.788	(0.760 , 0.817)
B&P's Kappa ($\hat{k}_n$) :	0.819	(0.795 , 0.844)
Tau (T_p) :	0.788	(0.759 , 0.817)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.0 Probability Threshold
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	523	0	53	0	0	576
road	0	170	9	0	0	179
grass	0	4	317	0	0	321
metal	0	2	0	26	0	28
water	0	0	2	0	98	100
	523	176	381	26	98	

Reference Source / # :	Dependent/Synthetic - 1204	
Pixels Verified :	1204	
Diagonal Elements :	1134	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.942	(0.929 , 0.955)
Weighted Accuracy (w) :	0.960	(0.939 , 0.980)
Kappa ($\hat{k}$) :	0.914	(0.895 , 0.934)
B&P's Kappa ($\hat{k}_n$) :	0.927	(0.911 , 0.944)
Tau (T_p) :	0.915	(0.895 , 0.934)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	519	108	222	1	0	850
road	0	67	2	0	0	69
grass	2	0	157	17	0	176
metal	2	1	0	8	1	12
water	0	0	0	0	97	97
	523	176	381	26	98	

Reference Source / # :	Dependent/Synthetic - 1204	
Pixels Verified :	1204	
Diagonal Elements :	848	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.704	(0.679 , 0.730)
Weighted Accuracy (w) :	0.617	(0.528 , 0.705)
Kappa ($\hat{k}$) :	0.532	(0.494 , 0.570)
B&P's Kappa ($\hat{k}_n$) :	0.630	(0.598 , 0.663)
Tau (T_p) :	0.567	(0.529 , 0.605)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	523	0	53	0	0	576
road	0	170	155	0	0	325
grass	0	6	169	26	0	201
metal	0	0	0	0	0	0
water	0	0	4	0	98	102
	523	176	381	26	98	

Reference Source / # :	Dependent/Synthetic - 1204
Pixels Verified :	1204
Diagonal Elements :	960

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.797	(0.775 , 0.820)
Weighted Accuracy (w) :	0.682	(0.657 , 0.707)
Kappa ($\hat{k}$) :	0.708	(0.677 , 0.738)
B&P's Kappa ($\hat{k}_n$) :	0.747	(0.718 , 0.775)
Tau (T_p) :	0.703	(0.670 , 0.736)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.0 Probability Threshold
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	523	0	53	0	0	576
road	0	170	31	0	0	201
grass	0	4	254	0	0	258
metal	0	1	0	24	0	25
water	0	1	43	2	98	144
	523	176	381	26	98	

Reference Source / # :	Dependent/Synthetic - 1204
Pixels Verified :	1204
Diagonal Elements :	1069

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.888	(0.870 , 0.906)
Weighted Accuracy (w) :	0.911	(0.859 , 0.963)
Kappa ($\hat{k}$) :	0.837	(0.812 , 0.863)
B&P's Kappa ($\hat{k}_n$) :	0.860	(0.838 , 0.882)
Tau (T_p) :	0.836	(0.810 , 0.862)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	ratio two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	502	2	26	1	0	531
road	0	166	9	0	0	175
grass	18	8	346	13	0	385
metal	3	0	0	12	0	15
water	0	0	0	0	98	98
	523	176	381	26	98	

Reference Source / # :	Dependent/Synthetic - 1204
Pixels Verified :	1204
Diagonal Elements :	1124

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.934	(0.919 , 0.948)
Weighted Accuracy (w) :	0.855	(0.766 , 0.943)
Kappa ($\hat{k}$) :	0.902	(0.881 , 0.923)
B&P's Kappa ($\hat{k}_n$) :	0.917	(0.899 , 0.935)
Tau (T_p) :	0.903	(0.882 , 0.923)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	522	10	52	0	0	584
road	1	161	114	0	0	276
grass	0	5	211	26	0	242
metal	0	0	0	0	0	0
water	0	0	4	0	98	102
	523	176	381	26	98	

Reference Source / # :	Dependent/Synthetic - 1204	
Pixels Verified :	1204	
Diagonal Elements :	992	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.824	(0.802 , 0.845)
Weighted Accuracy (w) :	0.693	(0.664 , 0.722)
Kappa ($\hat{k}$) :	0.743	(0.713 , 0.773)
B&P's Kappa ($\hat{k}_n$) :	0.780	(0.753 , 0.807)
Tau (T_p) :	0.742	(0.711 , 0.774)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.0 Probability Threshold
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	268	0	5	16	0	289
road	0	160	32	4	0	196
grass	0	40	1832	0	0	1872
metal	0	1	2	50	0	53
water	0	0	1	0	301	302
	268	201	1872	70	301	

Reference Source / # :	Independent/Synthetic - 2712	
Pixels Verified :	2712	
Diagonal Elements :	2611	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.963	(0.956 , 0.970)
Weighted Accuracy (w) :	0.898	(0.844 , 0.951)
Kappa ($\hat{k}$) :	0.925	(0.910 , 0.939)
B&P's Kappa ($\hat{k}_n$) :	0.953	(0.945 , 0.962)
Tau (T_p) :	0.925	(0.910 , 0.939)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	5km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	205	185	508	0	0	898
road	0	0	0	0	0	0
grass	4	14	1357	23	0	1398
metal	0	1	1	6	0	8
water	59	1	6	41	301	408
	268	201	1872	70	301	

Reference Source / # :	Independent/Synthetic - 2712
Pixels Verified :	2712
Diagonal Elements :	1869

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.689	(0.672 , 0.707)
Weighted Accuracy (w) :	0.515	(0.477 , 0.553)
Kappa ($\hat{k}$) :	0.477	(0.453 , 0.502)
B&P's Kappa ($\hat{k}_n$) :	0.611	(0.590 , 0.633)
Tau (T_p) :	0.373	(0.337 , 0.408)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 5km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	265	152	35	40	0	492
road	3	40	673	3	0	719
grass	0	9	1093	7	0	1109
metal	0	0	0	0	0	0
water	0	0	71	20	301	392
	268	201	1872	70	301	

Reference Source / # : Independent/Synthetic - 2712

Pixels Verified : 2712

Diagonal Elements : 1699

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.626	(0.608 , 0.645)
Weighted Accuracy (w) :	0.554	(0.527 , 0.582)
Kappa ($\hat{k}$) :	0.438	(0.413 , 0.462)
B&P's Kappa ($\hat{k}_n$) :	0.533	(0.510 , 0.556)
Tau (T_p) :	0.246	(0.209 , 0.283)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 7km Visibility

Classifier : Gaussian Maximum Likelihood

Notes : 0.0 Probability Threshold

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	268	1	2	2	0	273
road	0	169	20	0	0	189
grass	0	31	1841	0	0	1872
metal	0	0	8	68	9	85
water	0	0	1	0	292	293
	268	201	1872	70	301	

Reference Source / # : Independent/Synthetic - 2712

Pixels Verified : 2712

Diagonal Elements : 2638

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.973	(0.967 , 0.979)
Weighted Accuracy (w) :	0.953	(0.923 , 0.983)
Kappa ($\hat{k}$) :	0.945	(0.933 , 0.957)
B&P's Kappa ($\hat{k}_n$) :	0.966	(0.958 , 0.974)
Tau (T_p) :	0.945	(0.933 , 0.957)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	7km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	267	143	823	3	0	1236
road	0	57	0	0	0	57
grass	0	0	1042	0	0	1042
metal	1	1	7	67	0	76
water	0	0	0	0	301	301
	268	201	1872	70	301	

Reference Source / # :	Independent/Synthetic - 2712
Pixels Verified :	2712
Diagonal Elements :	1734

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.639	(0.621 , 0.657)
Weighted Accuracy (w) :	0.759	(0.722 , 0.795)
Kappa ($\hat{k}$) :	0.466	(0.442 , 0.490)
B&P's Kappa ($\hat{k}_n$) :	0.549	(0.527 , 0.572)
Tau (T_p) :	0.272	(0.235 , 0.308)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 7km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	268	2	6	5	0	281
road	0	193	852	0	0	1045
grass	0	6	829	32	0	867
metal	0	0	0	0	0	0
water	0	0	185	33	301	519
	268	201	1872	70	301	

Reference Source / # : Independent/Synthetic - 2712

Pixels Verified : 2712

Diagonal Elements : 1591

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.587	(0.568 , 0.605)
Weighted Accuracy (w) :	0.681	(0.665 , 0.696)
Kappa ($\hat{k}$) :	0.425	(0.402 , 0.449)
B&P's Kappa ($\hat{k}_n$) :	0.483	(0.460 , 0.506)
Tau (T_p) :	0.165	(0.128 , 0.203)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Gaussian Maximum Likelihood

Notes : 0.0 Probability Threshold

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	268	0	3	0	45	316
road	0	157	29	0	0	186
grass	0	43	1663	0	0	1706
metal	0	1	5	69	0	75
water	0	0	172	1	256	429
	268	201	1872	70	301	

Reference Source / # : Independent/Synthetic - 2712

Pixels Verified : 2712

Diagonal Elements : 2413

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.890	(0.878 , 0.902)
Weighted Accuracy (w) :	0.901	(0.867 , 0.935)
Kappa ($\hat{k}$) :	0.792	(0.770 , 0.814)
B&P's Kappa ($\hat{k}_n$) :	0.862	(0.847 , 0.877)
Tau (T_p) :	0.777	(0.754 , 0.801)

Scene Name :	Forest Scene
Image Dimensions :	250 x 250 x 15 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	ratio two bands, large training set
Total Pixels :	62500
Pixels Classified :	62500
Percent Classified :	100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	264	0	0	0	0	264
road	0	159	11	0	0	170
grass	2	42	1857	50	15	1966
metal	2	0	4	20	0	26
water	0	0	0	0	286	286
	268	201	1872	70	301	

Reference Source / # :	Independent/Synthetic - 2712
Pixels Verified :	2712
Diagonal Elements :	2586

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.954	(0.946 , 0.961)
Weighted Accuracy (w) :	0.801	(0.746 , 0.856)
Kappa ($\hat{k}$) :	0.902	(0.885 , 0.919)
B&P's Kappa ($\hat{k}_n$) :	0.942	(0.932 , 0.952)
Tau (T_p) :	0.906	(0.890 , 0.922)

Scene Name : Forest Scene

Image Dimensions : 250 x 250 x 15 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 62500

Pixels Classified : 62500

Percent Classified : 100.0

Confusion Matrix

	forest	road	grass	metal	water	
forest	266	16	28	0	0	310
road	1	169	373	0	0	543
grass	1	16	1466	41	0	1524
metal	0	0	0	0	0	0
water	0	0	5	29	301	335
	268	201	1872	70	301	

Reference Source / # : Independent/Synthetic - 2712

Pixels Verified : 2717

Diagonal Elements : 2202

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.812	(0.797 , 0.827)
Weighted Accuracy (w) :	0.723	(0.699 , 0.748)
Kappa ($\hat{k}$) :	0.671	(0.646 , 0.696)
B&P's Kappa ($\hat{k}_n$) :	0.765	(0.747 , 0.783)
Tau (T_p) :	0.620	(0.591 , 0.650)

Scene Name : Desert Scene

Image Dimensions : 600 x 2000 x 10 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Gaussian Maximum Likelihood

Notes : 0.15 Probability Threshold

Total Pixels : 1200000

Pixels Classified : 1200000

Percent Classified : 100.0

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	4265	0	0	0	0	0	4265
sand	0	1616	0	0	0	0	1616
tarp	0	5	1785	0	3	0	1793
vegetation	0	0	0	937	2	49	988
shadow	0	0	2	1	382	2	387
camouflage	0	0	4	38	6	458	506
	4265	1621	1791	976	393	509	

Reference Source / # : Dependent - 9555

Pixels Verified : 9555

Diagonal Elements : 9443

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.988	(0.986 , 0.990)
Weighted Accuracy (w) :	0.971	(0.965 , 0.976)
Kappa ($\hat{k}$) :	0.984	(0.981 , 0.987)
B&P's Kappa ($\hat{k}_n$) :	0.986	(0.983 , 0.989)
Tau (T_p) :	0.984	(0.981 , 0.987)

Scene Name :	Desert Scene
Image Dimensions :	600 x 2000 x 10 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold_3_ratio_1pair rule with large initial training set
Total Pixels :	1200000
Pixels Classified :	1071910
Percent Classified :	89.3

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	4213	0	0	13	4	3	4233
sand	0	1621	0	0	0	0	1621
tarp	1	0	1684	0	0	0	1685
vegetation	0	0	0	327	0	0	327
shadow	0	0	99	0	209	0	308
camouflage	50	0	8	636	180	505	1379
	4264	1621	1791	976	393	508	

Reference Source / # :	Dependent - 9555	
Pixels Verified :	9553	
Diagonal Elements :	8559	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.896	(0.890 , 0.902)
Weighted Accuracy (w) :	0.798	(0.788 , 0.808)
Kappa ($\hat{k}$) :	0.857	(0.849 , 0.865)
B&P's Kappa ($\hat{k}_n$) :	0.875	(0.868 , 0.882)
Tau (T_p) :	0.856	(0.847 , 0.864)

Scene Name :	Desert Scene
Image Dimensions :	600 x 2000 x 10 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	1200000
Pixels Classified :	1199110
Percent Classified :	99.9

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	4260	0	0	0	0	4	4264
sand	0	1621	0	0	0	0	1621
tarp	0	0	1791	1	4	1	1797
vegetation	0	0	0	844	26	158	1028
shadow	0	0	0	8	362	3	373
camouflage	5	0	0	123	1	343	472
	4265	1621	1791	976	393	509	

Reference Source / # :	Dependent - 9555	
Pixels Verified :	9555	
Diagonal Elements :	9222	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.965	(0.961 , 0.969)
Weighted Accuracy (w) :	0.910	(0.901 , 0.919)
Kappa ($\hat{k}$) :	0.952	(0.947 , 0.957)
B&P's Kappa ($\hat{k}_n$) :	0.958	(0.954 , 0.962)
Tau (T_p) :	0.952	(0.946 , 0.957)

Scene Name :	Desert Scene
Image Dimensions :	300 x 1000 x 10 bands
Resolution :	2m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	300000
Pixels Classified :	300000
Percent Classified :	100.0

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	555	0	0	0	0	0	555
sand	0	423	0	0	0	0	423
tarp	0	0	358	0	0	0	358
vegetation	0	0	0	357	0	0	357
shadow	0	0	0	0	146	0	146
camouflage	0	0	0	0	0	232	232
	555	423	358	357	146	232	

Reference Source / # :	Dependent - 2071	
Pixels Verified :	2071	
Diagonal Elements :	2071	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	1.0	(1.0 , 1.0)
Weighted Accuracy (w) :	1.0	(1.0 , 1.0)
Kappa ($\hat{k}$) :	1.0	(1.0 ,1.0)
B&P's Kappa ($\hat{k}_n$) :	1.0	(1.0 , 1.0)
Tau (T_p) :	1.0	(1.0 , 1.0)

Scene Name : Desert Scene

Image Dimensions : 300 x 1000 x 10 bands

Resolution : 2m GIFOV

Atmosphere : 23km Visibility

Classifier : Rule Based Genetic Algorithm (MYSTIC)

Notes : threshold_3_ratio_1pair rule with large initial training set

Total Pixels : 300000

Pixels Classified :269581

Percent Classified : 89.9

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	482	0	0	2	5	0	489
sand	0	423	0	0	0	0	423
tarp	0	0	358	0	0	0	358
vegetation	0	0	0	121	0	3	124
shadow	0	0	0	0	74	0	74
camouflage	73	0	0	234	66	229	602
	555	423	358	357	145	232	

Reference Source / # : Dependent - 2071

Pixels Verified : 2070

Diagonal Elements : 1687

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.815	(0.798 , 0.832)
Weighted Accuracy (w) :	0.784	(0.767 , 0.801)
Kappa ($\hat{k}$) :	0.774	(0.755 , 0.794)
B&P's Kappa ($\hat{k}_n$) :	0.778	(0.758 , 0.798)
Tau (T_p) :	0.771	(0.751 , 0.792)

Scene Name : Desert Scene

Image Dimensions : 300 x 1000 x 10 bands

Resolution : 2m GIFOV

Atmosphere : 23km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 300000

Pixels Classified :298586

Percent Classified : 99.5

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	555	0	0	0	0	0	555
sand	0	423	0	0	0	0	423
tarp	0	0	358	0	0	0	358
vegetation	0	0	0	301	0	46	347
shadow	0	0	0	0	138	1	139
camouflage	0	0	0	56	8	185	249
	555	423	358	357	146	232	

Reference Source / # : Dependent - 2071

Pixels Verified : 2071

Diagonal Elements : 1960

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.946	(0.937 , 0.956)
Weighted Accuracy (w) :	0.931	(0.919 , 0.943)
Kappa ($\hat{k}$) :	0.934	(0.922 , 0.946)
B&P's Kappa ($\hat{k}_n$) :	0.936	(0.924 , 0.947)
Tau (T_p) :	0.934	(0.922 , 0.946)

Scene Name :	Desert Scene
Image Dimensions :	150 x 500 x 10 bands
Resolution :	4m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	75000
Pixels Classified :	75000
Percent Classified :	100.0

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	192	0	0	0	0	0	192
sand	0	191	0	0	0	0	191
tarp	0	0	133	0	0	0	133
vegetation	0	0	0	98	0	0	98
shadow	0	0	0	0	100	0	100
camouflage	0	0	0	0	0	96	96
	192	191	133	98	100	96	

Reference Source / # :	Dependent - 810	
Pixels Verified :	810	
Diagonal Elements :	810	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	1.0	(1.0 , 1.0)
Weighted Accuracy (w) :	1.0	(1.0 , 1.0)
Kappa ($\hat{k}$) :	1.0	(1.0 , 1.0)
B&P's Kappa ($\hat{k}_n$) :	1.0	(1.0 , 1.0)
Tau (T_p) :	1.0	(1.0 , 1.0)

Scene Name :	Desert Scene
Image Dimensions :	150 x 500 x 10 bands
Resolution :	4m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold_3_ratio_1pair rule with large initial training set
Total Pixels :	75000
Pixels Classified :	67140
Percent Classified :	89.5

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	169	0	0	3	24	0	196
sand	0	165	0	0	0	0	165
tarp	0	1	113	0	0	0	114
vegetation	0	0	0	24	1	1	26
shadow	0	0	14	0	9	1	24
camouflage	23	0	6	70	64	94	257
	192	166	133	97	98	96	

Reference Source / # :	Dependent - 810	
Pixels Verified :	782	
Diagonal Elements :	574	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.734	(0.703 , 0.765)
Weighted Accuracy (w) :	0.674	(0.652 , 0.696)
Kappa ($\hat{k}$) :	0.676	(0.641 , 0.711)
B&P's Kappa ($\hat{k}_n$) :	0.681	(0.644 , 0.718)
Tau (T_p) :	0.675	(0.638 , 0.713)

Scene Name : Desert Scene

Image Dimensions : 150 x 500 x 10 bands

Resolution : 4m GIFOV

Atmosphere : 23km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 75000

Pixels Classified :74945

Percent Classified : 99.9

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	186	0	0	0	0	0	186
sand	0	191	0	0	0	0	191
tarp	0	0	133	1	5	0	139
vegetation	0	0	0	82	0	6	88
shadow	6	0	0	10	94	23	133
camouflage	0	0	0	5	1	67	73
	192	191	133	98	100	96	

Reference Source / # : Dependent - 810

Pixels Verified : 810

Diagonal Elements : 753

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.930	(0.912 , 0.947)
Weighted Accuracy (w) :	0.907	(0.886 , 0.929)
Kappa ($\hat{k}$) :	0.914	(0.893 , 0.935)
B&P's Kappa ($\hat{k}_n$) :	0.916	(0.894 , 0.937)
Tau (T_p) :	0.914	(0.892 , 0.935)

Scene Name :	Desert Scene
Image Dimensions :	600 x 2000 x 10 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	1200000
Pixels Classified :	1200000
Percent Classified :	100.0

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	7238	2	0	0	0	0	7240
sand	0	1236	0	0	0	0	1236
tarp	0	60	1651	1	24	10	1746
vegetation	0	0	0	877	31	32	940
shadow	0	0	0	0	876	23	899
camouflage	0	2	1	206	30	664	903
	7238	1300	1652	1084	961	729	

Reference Source / # :	Independent - 12964	
Pixels Verified :	12964	
Diagonal Elements :	12542	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.967	(0.964 , 0.971)
Weighted Accuracy (w) :	0.930	(0.924 , 0.937)
Kappa ($\hat{k}$) :	0.950	(0.945 , 0.954)
B&P's Kappa ($\hat{k}_n$) :	0.961	(0.957 , 0.965)
Tau (T_p) :	0.950	(0.945 , 0.954)

Scene Name : Desert Scene

Image Dimensions : 600 x 2000 x 10 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Rule Based Genetic Algorithm (MYSTIC)

Notes : threshold_3_ratio_1pair rule with large initial training set

Total Pixels : 1200000

Pixels Classified : 1071910

Percent Classified : 89.3

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	7209	6	0	45	27	9	7296
sand	0	1251	12	0	0	0	1263
tarp	0	6	1560	0	0	0	1566
vegetation	0	0	0	264	2	4	270
shadow	0	0	0	0	195	0	195
camouflage	29	3	36	774	737	716	2295
	7238	1266	1608	1083	961	729	

Reference Source / # : Independent - 12964

Pixels Verified : 12885

Diagonal Elements : 11195

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.869	(0.863 , 0.875)
Weighted Accuracy (w) :	0.731	(0.724 , 0.737)
Kappa ($\hat{k}$) :	0.796	(0.788 , 0.804)
B&P's Kappa ($\hat{k}_n$) :	0.843	(0.836 , 0.850)
Tau (T_p) :	0.796	(0.787 , 0.805)

Scene Name : Desert Scene

Image Dimensions : 600 x 2000 x 10 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 1200000

Pixels Classified : 1199110

Percent Classified : 99.9

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	7234	9	3	5	4	67	7322
sand	0	1291	97	0	0	0	1388
tarp	0	0	863	8	23	12	906
vegetation	0	0	0	721	81	186	988
shadow	0	0	0	40	827	24	891
camouflage	4	0	0	310	26	440	780
	7238	1300	963	1084	961	729	

Reference Source / # : Independent - 12964

Pixels Verified : 12275

Diagonal Elements : 11376

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.927	(0.922 , 0.931)
Weighted Accuracy (w) :	0.836	(0.827 , 0.845)
Kappa ($\hat{k}$) :	0.881	(0.874 , 0.888)
B&P's Kappa ($\hat{k}_n$) :	0.912	(0.907 , 0.918)
Tau (T_p) :	0.881	(0.874 , 0.889)

Scene Name :	Desert Scene
Image Dimensions :	300 x 1000 x 10 bands
Resolution :	2m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	300000
Pixels Classified :	300000
Percent Classified :	100.0

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	1854	103	23	3	0	3	1986
sand	0	383	0	0	0	0	383
tarp	0	0	362	0	5	8	375
vegetation	0	0	0	231	17	2	250
shadow	0	1	0	0	180	1	182
camouflage	0	0	0	48	105	248	401
	1854	487	385	282	307	262	

Reference Source / # :	Independent - 3577
Pixels Verified :	3577
Diagonal Elements :	3258

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.911	(0.901 , 0.920)
Weighted Accuracy (w) :	0.846	(0.832 , 0.861)
Kappa ($\hat{k}$) :	0.867	(0.853 , 0.880)
B&P's Kappa ($\hat{k}_n$) :	0.893	(0.882 , 0.904)
Tau (T_p) :	0.869	(0.856 , 0.883)

Scene Name :	Desert Scene
Image Dimensions :	300 x 1000 x 10 bands
Resolution :	2m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold_3_ratio_1pair rule with large initial training set
Total Pixels :	300000
Pixels Classified :	269581
Percent Classified :	89.9

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	1721	0	0	6	15	8	1750
sand	0	448	0	0	0	0	448
tarp	0	2	358	0	0	1	361
vegetation	0	0	0	55	2	22	79
shadow	0	0	21	0	45	1	67
camouflage	133	0	4	221	240	225	823
	1854	450	383	282	302	257	

Reference Source / # :	Independent - 3577	
Pixels Verified :	3528	
Diagonal Elements :	2852	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.808	(0.795 , 0.821)
Weighted Accuracy (w) :	0.680	(0.667 , 0.693)
Kappa ($\hat{k}$) :	0.723	(0.706 , 0.740)
B&P's Kappa ($\hat{k}_n$) :	0.770	(0.754 , 0.786)
Tau (T_p) :	0.717	(0.698 , 0.736)

Scene Name : Desert Scene

Image Dimensions : 300 x 1000 x 10 bands

Resolution : 2m GIFOV

Atmosphere : 23km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 300000

Pixels Classified :298586

Percent Classified : 99.5

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	1842	0	10	4	4	11	1871
sand	0	464	0	0	0	0	464
tarp	0	3	327	0	0	0	330
vegetation	0	0	0	199	18	73	290
shadow	10	20	40	2	253	11	336
camouflage	2	0	0	77	32	167	278
	1854	487	377	282	307	262	

Reference Source / # : Independent - 3577

Pixels Verified : 3569

Diagonal Elements : 3252

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.911	(0.902 , 0.921)
Weighted Accuracy (w) :	0.830	(0.814 , 0.846)
Kappa ($\hat{k}$) :	0.869	(0.856 , 0.883)
B&P's Kappa ($\hat{k}_n$) :	0.893	(0.882 , 0.905)
Tau (T_p) :	0.870	(0.856 , 0.883)

Scene Name : Desert Scene

Image Dimensions : 150 x 500 x 10 bands

Resolution : 4m GIFOV

Atmosphere : 23km Visibility

Classifier : Gaussian Maximum Likelihood

Notes : 0.15 Probability Threshold

Total Pixels : 75000

Pixels Classified : 75000

Percent Classified : 100.0

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	339	0	0	0	0	0	339
sand	0	140	0	0	0	1	141
tarp	0	8	118	2	12	10	150
vegetation	0	0	0	101	1	2	104
shadow	8	0	0	24	107	36	175
camouflage	0	0	0	46	6	85	137
	347	148	118	173	126	134	

Reference Source / # : Independent - 1046

Pixels Verified : 1046

Diagonal Elements : 890

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.851	(0.829 , 0.872)
Weighted Accuracy (w) :	0.832	(0.810 , 0.854)
Kappa ($\hat{k}$) :	0.814	(0.788 , 0.841)
B&P's Kappa ($\hat{k}_n$) :	0.821	(0.795 , 0.847)
Tau (T_p) :	0.813	(0.786 , 0.840)

Scene Name :	Desert Scene
Image Dimensions :	150 x 500 x 10 bands
Resolution :	4m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold_3_ratio_1pair rule with large initial training set
Total Pixels :	75000
Pixels Classified :	67140
Percent Classified :	89.5

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	310	2	0	6	14	7	339
sand	0	137	0	0	0	1	138
tarp	0	1	113	0	0	0	114
vegetation	0	0	0	21	0	14	35
shadow	0	0	1	0	35	0	36
camouflage	37	0	2	144	72	110	365
	347	140	116	171	121	132	

Reference Source / # :	Independent - 1046
Pixels Verified :	1027
Diagonal Elements :	726

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.707	(0.679 , 0.735)
Weighted Accuracy (w) :	0.682	(0.661 , 0.703)
Kappa ($\hat{k}$) :	0.635	(0.602 , 0.667)
B&P's Kappa ($\hat{k}_n$) :	0.648	(0.615 , 0.682)
Tau (T_p) :	0.632	(0.597 , 0.667)

Scene Name : Desert Scene

Image Dimensions : 150 x 500 x 10 bands

Resolution : 4m GIFOV

Atmosphere : 23km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 75000

Pixels Classified :74945

Percent Classified : 99.9

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	267	2	0	1	0	1	271
sand	0	143	1	0	0	1	145
tarp	0	0	53	4	27	5	89
vegetation	0	0	0	101	1	19	121
shadow	80	3	13	41	89	62	288
camouflage	0	0	0	26	9	46	81
	347	148	67	173	126	134	

Reference Source / # : Independent - 1046

Pixels Verified : 995

Diagonal Elements : 699

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.703	(0.674 , 0.731)
Weighted Accuracy (w) :	0.693	(0.664 , 0.722)
Kappa ($\hat{k}$) :	0.632	(0.598 , 0.666)
B&P's Kappa ($\hat{k}_n$) :	0.643	(0.609 , 0.677)
Tau (T_p) :	0.622	(0.586 , 0.658)

Scene Name : Desert Scene

Image Dimensions : 600 x 2000 x 10 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Gaussian Maximum Likelihood

Notes : 0.15 Probability Threshold

Total Pixels : 1200000

Pixels Classified : 1200000

Percent Classified : 100.0

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	10499	1	0	0	0	0	10500
sand	0	804	0	0	0	0	804
tarp	0	39	425	0	5	7	476
vegetation	0	0	0	415	7	21	443
shadow	0	0	0	0	187	15	202
camouflage	0	1	0	97	6	435	539
	10499	845	425	512	205	478	

Reference Source / # : Scaled Independent - 12964

Pixels Verified : 12964

Diagonal Elements : 12765

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.985	(0.983 , 0.987)
Weighted Accuracy (w) :	0.931	(0.906 , 0.955)
Kappa ($\hat{k}$) :	0.954	(0.948 , 0.960)
B&P's Kappa ($\hat{k}_n$) :	0.982	(0.979 , 0.984)
Tau (T_p) :	0.954	(0.948 , 0.961)

Scene Name :	Desert Scene
Image Dimensions :	600 x 2000 x 10 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold_3_ratio_1pair rule with large initial training set
Total Pixels :	1200000
Pixels Classified :	1071910
Percent Classified :	89.3

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	10940	3	0	2	2	13	10960
sand	0	532	1	0	0	0	533
tarp	0	3	160	0	0	0	163
vegetation	0	0	0	10	0	6	16
shadow	0	0	0	0	14	0	14
camouflage	44	1	4	29	51	1071	1200
	10984	539	165	41	67	1090	

Reference Source / # :	Scaled Independent - 12964	
Pixels Verified :	12886	
Diagonal Elements :	12727	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.988	(0.863 , 0.875)
Weighted Accuracy (w) :	0.731	(0.665 , 0.799)
Kappa ($\hat{k}$) :	0.953	(0.946 , 0.961)
B&P's Kappa ($\hat{k}_n$) :	0.985	(0.983 , 0.987)
Tau (T_p) :	0.953	(0.946 , 0.961)

Scene Name : Desert Scene

Image Dimensions : 600 x 2000 x 10 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 1200000

Pixels Classified : 1199110

Percent Classified : 99.9

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	9795	7	0	1	4	16	9823
sand	0	966	14	0	0	0	980
tarp	0	0	126	1	24	3	154
vegetation	0	0	0	114	86	44	244
shadow	0	0	0	6	873	6	885
camouflage	5	0	0	49	27	105	186
	9800	973	140	171	1014	174	

Reference Source / # : Scaled Independent - 12964

Pixels Verified : 12272

Diagonal Elements : 11979

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.976	(0.973 , 0.979)
Weighted Accuracy (w) :	0.837	(0.790 , 0.884)
Kappa ($\hat{k}$) :	0.931	(0.924 , 0.939)
B&P's Kappa ($\hat{k}_n$) :	0.971	(0.968 , 0.975)
Tau (T_p) :	0.932	(0.924 , 0.939)

Scene Name : Desert Scene

Image Dimensions : 300 x 1000 x 10 bands

Resolution : 2m GIFOV

Atmosphere : 23km Visibility

Classifier : Gaussian Maximum Likelihood

Notes : 0.15 Probability Threshold

Total Pixels : 300000

Pixels Classified : 300000

Percent Classified : 100.0

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	3135	20	1	1	0	1	3158
sand	0	75	0	0	0	0	75
tarp	0	0	14	0	2	3	19
vegetation	0	0	0	108	6	1	115
shadow	0	0	0	0	66	0	66
camouflage	0	0	0	22	39	83	144
	3135	95	15	131	113	88	

Reference Source / # : Scaled Independent - 3577

Pixels Verified : 3577

Diagonal Elements : 3481

	Coefficients	95% Confidence Interval
Simple Accuracy (P_0) :	0.973	(0.968 , 0.978)
Weighted Accuracy (w) :	0.846	(0.767 , 0.925)
Kappa ($\hat{k}$) :	0.880	(0.858 , 0.901)
B&P's Kappa ($\hat{k}_n$) :	0.968	(0.961 , 0.974)
Tau (T_p) :	0.882	(0.859 , 0.906)

Scene Name :	Desert Scene
Image Dimensions :	300 x 1000 x 10 bands
Resolution :	2m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold_3_ratio_1pair rule with large initial training set
Total Pixels :	300000
Pixels Classified :	269581
Percent Classified :	89.9

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	2728	0	0	0	1	13	2742
sand	0	137	0	0	0	0	137
tarp	0	1	23	0	0	2	26
vegetation	0	0	0	2	0	35	37
shadow	0	0	1	0	2	2	5
camouflage	211	0	0	7	10	355	583
	2939	138	24	9	13	407	

Reference Source / # :	Scaled Independent - 3577
Pixels Verified :	3530
Diagonal Elements :	3247

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.920	(0.795 , 0.821)
Weighted Accuracy (w) :	0.688	(0.547 , 0.829)
Kappa ($\hat{k}$) :	0.759	(0.733 , 0.785)
B&P's Kappa ($\hat{k}_n$) :	0.904	(0.893 , 0.915)
Tau (T_p) :	0.725	(0.695 , 0.756)

Scene Name :	Desert Scene
Image Dimensions :	300 x 1000 x 10 bands
Resolution :	2m GIFOV
Atmosphere :	23km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	300000
Pixels Classified :	298586
Percent Classified :	99.5

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	2661	0	1	1	7	4	2674
sand	0	145	0	0	0	0	145
tarp	0	1	18	0	0	0	19
vegetation	0	0	0	40	33	26	99
shadow	14	6	2	0	467	4	493
camouflage	3	0	0	16	59	60	138
	2678	152	21	57	566	94	

Reference Source / # :	Scaled Independent - 3577
Pixels Verified :	3568
Diagonal Elements :	3391

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.950	(0.943 , 0.958)
Weighted Accuracy (w) :	0.828	(0.739 , 0.918)
Kappa ($\hat{k}$) :	0.880	(0.864 , 0.896)
B&P's Kappa ($\hat{k}_n$) :	0.940	(0.932 , 0.949)
Tau (T_p) :	0.879	(0.861 , 0.896)

Scene Name :	Desert Scene
Image Dimensions :	150 x 500 x 10 bands
Resolution :	4m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	75000
Pixels Classified :	75000
Percent Classified :	100.0

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	730	0	0	0	0	0	730
sand	0	91	0	0	0	0	91
tarp	0	5	50	0	11	1	67
vegetation	0	0	0	15	1	0	16
shadow	17	0	0	3	97	4	121
camouflage	0	0	0	7	5	9	21
	747	96	50	25	114	14	

Reference Source / # :	Scaled Independent - 1046
Pixels Verified :	1046
Diagonal Elements :	992

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.948	(0.935 , 0.962)
Weighted Accuracy (w) :	0.836	(0.703 , 0.970)
Kappa ($\hat{k}$) :	0.892	(0.865 , 0.919)
B&P's Kappa ($\hat{k}_n$) :	0.938	(0.922 , 0.954)
Tau (T_p) :	0.889	(0.861 , 0.918)

Scene Name :	Desert Scene
Image Dimensions :	150 x 500 x 10 bands
Resolution :	4m GIFOV
Atmosphere :	23km Visibility

Classifier :	Rule Based Genetic Algorithm (MYSTIC)
Notes :	threshold_3_ratio_1pair rule with large initial training set
Total Pixels :	75000
Pixels Classified :	67140
Percent Classified :	89.5

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	776	1	0	0	0	6	783
sand	0	34	0	0	0	1	35
tarp	0	0	5	0	0	0	5
vegetation	0	0	0	0	0	12	12
shadow	0	0	0	0	1	0	1
camouflage	93	0	0	2	1	95	191
	869	35	5	2	2	114	

Reference Source / # :	Scaled Independent - 1046	
Pixels Verified :	1027	
Diagonal Elements :	911	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.887	(0.868 , 0.906)
Weighted Accuracy (w) :	0.700	(0.414 , 0.985)
Kappa ($\hat{k}$) :	0.661	(0.606 , 0.715)
B&P's Kappa ($\hat{k}_n$) :	0.864	(0.841 , 0.888)
Tau (T_p) :	0.582	(0.511 , 0.654)

Scene Name :	Desert Scene
Image Dimensions :	150 x 500 x 10 bands
Resolution :	4m GIFOV
Atmosphere :	23km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	75000
Pixels Classified :	74945
Percent Classified :	99.9

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	515	1	0	0	0	0	516
sand	0	66	0	0	0	0	66
tarp	0	0	3	0	49	0	52
vegetation	0	0	0	9	2	1	12
shadow	154	1	1	4	160	5	325
camouflage	0	0	0	2	16	4	22
	669	68	4	15	227	10	

Reference Source / # :	Scaled Independent - 1046	
Pixels Verified :	993	
Diagonal Elements :	757	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.762	(0.674 , 0.731)
Weighted Accuracy (w) :	0.699	(0.461 , 0.937)
Kappa ($\hat{k}$) :	0.583	(0.540 , 0.626)
B&P's Kappa ($\hat{k}_n$) :	0.715	(0.683 , 0.747)
Tau (T_p) :	0.514	(0.460 , 0.568)

Scene Name :	Desert Scene
Image Dimensions :	600 x 2000 x 10 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	1200000
Pixels Classified :	1200000
Percent Classified :	100.0

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	211	1	0	0	1	0	213
sand	9	25	0	0	0	0	34
tarp	7	2	3	0	1	0	13
vegetation	2	0	0	7	9	0	18
shadow	2	0	1	0	2	0	5
camouflage	3	0	0	0	3	8	14
	234	28	4	7	16	8	

Reference Source / # :	Random - 297
Pixels Verified :	297
Diagonal Elements :	256

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.862	(0.823 , 0.901)
Weighted Accuracy (w) :	0.778	(0.700 , 0.857)
Kappa ($\hat{k}$) :	0.671	(0.589 , 0.754)
B&P's Kappa ($\hat{k}_n$) :	0.834	(0.787 , 0.881)
Tau (T_p) :	0.623	(0.516 , 0.730)

Scene Name : Desert Scene

Image Dimensions : 600 x 2000 x 10 bands

Resolution : 1m GIFOV

Atmosphere : 23km Visibility

Classifier : Rule Based Genetic Algorithm (MYSTIC)

Notes : threshold_3_ratio_1pair rule with large initial training set

Total Pixels : 1200000

Pixels Classified : 1071910

Percent Classified : 89.3

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	198	1	0	1	3	1	204
sand	0	19	0	0	0	0	19
tarp	1	2	3	0	0	0	6
vegetation	0	0	0	0	1	0	1
shadow	0	0	0	0	1	0	1
camouflage	10	0	1	6	11	7	35
	209	22	4	7	16	8	

Reference Source / # : Random - 297

Pixels Verified : 266

Diagonal Elements : 228

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.857	(0.815 , 0.899)
Weighted Accuracy (w) :	0.583	(0.497 , 0.669)
Kappa ($\hat{k}$) :	0.631	(0.543 , 0.718)
B&P's Kappa ($\hat{k}_n$) :	0.829	(0.778 , 0.879)
Tau (T_p) :	0.614	(0.501 , 0.728)

Scene Name :	Desert Scene
Image Dimensions :	600 x 2000 x 10 bands
Resolution :	1m GIFOV
Atmosphere :	23km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	1200000
Pixels Classified :	1199110
Percent Classified :	99.9

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	200	1	0	0	1	1	203
sand	10	27	0	0	0	0	37
tarp	3	0	3	1	1	0	8
vegetation	1	0	0	2	3	1	7
shadow	0	0	1	2	5	0	8
camouflage	20	0	0	2	6	6	34
	234	28	4	7	16	8	

Reference Source / # :	Random - 297	
Pixels Verified :	297	
Diagonal Elements :	243	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.818	(0.774 , 0.862)
Weighted Accuracy (w) :	0.653	(0.542 , 0.763)
Kappa ($\hat{k}$) :	0.591	(0.505 , 0.676)
B&P's Kappa ($\hat{k}_n$) :	0.782	(0.729 , 0.834)
Tau (T_p) :	0.503	(0.383 , 0.623)

Scene Name :	Desert Scene
Image Dimensions :	300 x 1000 x 10 bands
Resolution :	2m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	300000
Pixels Classified :	300000
Percent Classified :	100.0

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	225	21	0	0	1	0	247
sand	2	7	0	0	0	0	9
tarp	0	0	3	0	0	0	3
vegetation	2	0	0	6	9	0	17
shadow	5	0	0	0	2	0	7
camouflage	0	0	1	1	4	8	14
	234	28	4	7	16	8	

Reference Source / # :	Random - 297	
Pixels Verified :	297	
Diagonal Elements :	251	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.845	(0.804 , 0.886)
Weighted Accuracy (w) :	0.657	(0.566 , 0.749)
Kappa ($\hat{k}$) :	0.542	(0.441 , 0.643)
B&P's Kappa ($\hat{k}_n$) :	0.814	(0.765 , 0.864)
Tau (T_p) :	0.577	(0.464 , 0.689)

Scene Name : Desert Scene

Image Dimensions : 300 x 1000 x 10 bands

Resolution : 2m GIFOV

Atmosphere : 23km Visibility

Classifier : Rule Based Genetic Algorithm (MYSTIC)

Notes : threshold_3_ratio_1pair rule with large initial training set

Total Pixels : 300000

Pixels Classified :269581

Percent Classified : 89.9

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	195	4	0	1	2	0	202
sand	1	16	0	0	0	0	17
tarp	0	1	3	0	0	0	4
vegetation	0	0	0	0	2	0	2
shadow	0	0	0	0	1	0	1
camouflage	12	1	1	6	11	8	39
	208	22	4	7	16	8	

Reference Source / # : Random - 297

Pixels Verified : 265

Diagonal Elements : 223

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.842	(0.798 , 0.885)
Weighted Accuracy (w) :	0.580	(0.500 , 0.659)
Kappa ($\hat{k}$) :	0.595	(0.506 , 0.684)
B&P's Kappa ($\hat{k}_n$) :	0.810	(0.757 , 0.863)
Tau (T_p) :	0.573	(0.455 , 0.692)

Scene Name : Desert Scene

Image Dimensions : 300 x 1000 x 10 bands

Resolution : 2m GIFOV

Atmosphere : 23km Visibility

Classifier : Fuzzy ARTMAP Neural Network

Notes : Recast Inconsistent Cases

Total Pixels : 300000

Pixels Classified : 298586

Percent Classified : 99.5

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	193	5	0	1	0	0	199
sand	1	20	0	0	0	0	21
tarp	0	0	3	0	0	0	3
vegetation	0	0	1	5	7	3	16
shadow	36	3	0	0	5	0	44
camouflage	2	0	0	1	4	5	12
	232	28	4	7	16	8	

Reference Source / # : Random - 297

Pixels Verified : 295

Diagonal Elements : 231

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.783	(0.736 , 0.830)
Weighted Accuracy (w) :	0.658	(0.542 , 0.774)
Kappa ($\hat{k}$) :	0.520	(0.431 , 0.609)
B&P's Kappa ($\hat{k}_n$) :	0.740	(0.683 , 0.796)
Tau (T_p) :	0.411	(0.283 , 0.538)

Scene Name :	Desert Scene
Image Dimensions :	150 x 500 x 10 bands
Resolution :	4m GIFOV
Atmosphere :	23km Visibility

Classifier :	Gaussian Maximum Likelihood
Notes :	0.15 Probability Threshold
Total Pixels :	75000
Pixels Classified :	75000
Percent Classified :	100.0

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	192	1	0	0	0	0	193
sand	20	20	0	0	0	0	40
tarp	5	4	3	1	1	0	14
vegetation	1	0	0	6	2	0	9
shadow	16	3	0	0	7	2	28
camouflage	0	0	1	0	6	6	13
	234	28	4	7	16	8	

Reference Source / # :	Random - 297
Pixels Verified :	297
Diagonal Elements :	234

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.788	(0.741 , 0.834)
Weighted Accuracy (w) :	0.722	(0.613 , 0.830)
Kappa ($\hat{k}$) :	0.546	(0.461 , 0.632)
B&P's Kappa ($\hat{k}_n$) :	0.745	(0.690 , 0.801)
Tau (T_p) :	0.420	(0.293 , 0.547)

Scene Name : Desert Scene

Image Dimensions : 150 x 500 x 10 bands

Resolution : 4m GIFOV

Atmosphere : 23km Visibility

Classifier : Rule Based Genetic Algorithm (MYSTIC)

Notes : threshold_3_ratio_1pair rule with large initial training set

Total Pixels : 75000

Pixels Classified :67140

Percent Classified : 89.5

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	196	4	0	1	1	1	203
sand	1	13	0	0	0	0	14
tarp	0	2	3	0	0	0	5
vegetation	0	0	0	0	2	0	2
shadow	0	0	0	0	1	0	1
camouflage	11	0	1	6	11	7	36
	208	19	4	7	15	8	

Reference Source / # : Random - 297

Pixels Verified : 261

Diagonal Elements : 220

	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.843	(0.799 , 0.887)
Weighted Accuracy (w) :	0.553	(0.463 , 0.643)
Kappa ($\hat{k}$) :	0.577	(0.485 , 0.668)
B&P's Kappa ($\hat{k}_n$) :	0.811	(0.759 , 0.864)
Tau (T_p) :	0.557	(0.432 , 0.681)

Scene Name :	Desert Scene
Image Dimensions :	150 x 500 x 10 bands
Resolution :	4m GIFOV
Atmosphere :	23km Visibility

Classifier :	Fuzzy ARTMAP Neural Network
Notes :	Recast Inconsistent Cases
Total Pixels :	75000
Pixels Classified :	74945
Percent Classified :	99.9

Confusion Matrix

	desert_pavement	sand	tarp	vegetation	shadow	camouflage	
desert_pavement	174	3	0	1	0	1	179
sand	11	18	0	0	0	0	29
tarp	0	0	4	0	0	0	4
vegetation	1	0	0	2	2	0	5
shadow	48	7	0	2	10	3	70
camouflage	0	0	0	2	4	4	10
	234	28	4	7	16	8	

Reference Source / # :	Random - 297	
Pixels Verified :	297	
Diagonal Elements :	212	
	Coefficients	95% Confidence Interval
Simple Accuracy (P_o) :	0.714	(0.662 , 0.765)
Weighted Accuracy (w) :	0.633	(0.538 , 0.728)
Kappa ($\hat{k}$) :	0.430	(0.344 , 0.515)
B&P's Kappa ($\hat{k}_n$) :	0.657	(0.595 , 0.718)
Tau (T_p) :	0.218	(0.0776 , 0.358)